

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 19, 2002, 06:53:36 ; Search time 25 Seconds

(without alignments)
709.342 Million cell updates/sec

Title: US-09-807-459-2

Perfect score: 2359

Sequence: 1 MAPSDVGDVTKTLAASES.....DPSKALIRKSTEDANLLEK 458

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	5.4	607	1	YMH1_YEAST
2	122	5.2	1829	1	FRPC_NEIMC
3	121	5.1	566	1	HEMA_IAUSS
4	120.5	5.1	566	1	HEMA_IAZUK
5	117	5.0	1038	1	CIN8_YEAST
6	117	5.0	1829	1	FRPC_NEIMB
7	116	4.9	437	1	Y162_UREPA
8	115	4.9	1875	1	MLP1_YEAST
9	113.5	4.8	2116	1	MYSL2_DICDI
10	113.5	4.8	2459	1	MAPB_RAT
11	112	4.7	566	1	HEMA_TALEN
12	112	4.7	1957	1	YD86_SCHPO
13	111.5	4.7	832	1	ALP6_SCHPO
14	110.5	4.7	886	1	ORC1_KLDUA
15	110.5	4.7	6359	1	BACQ_BACLI
16	110	4.7	566	1	HEMA_IAPUE
17	110	4.7	577	1	THPL_SCHPO
18	110	4.7	1271	1	Y338_MYCOE
19	108	4.6	566	1	HEMA_IATKE
20	107.5	4.6	462	1	YB47_METUA
21	107	4.5	550	1	HEMA_IATJ3
22	106.5	4.5	4036	1	RRPL_DUGBY
23	106	4.5	550	1	HEMA_IADH6
24	106	4.5	550	1	HEMA_IADH7
25	106	4.5	624	1	YMO8_YEAST
26	105.5	4.5	550	1	HEMA_IADHK
27	105.5	4.5	566	1	HEMA_IJEN7
28	105.5	4.5	1030	1	Y075_MYCPN
29	105.5	4.5	2670	1	YAO5_SCHPO
30	105	4.5	808	1	Y066_NPVAC
31	105	4.5	857	1	NFM_CHICK
32	104.5	4.4	778	1	YF05_METUA
33	104	4.4	550	1	HEMA_IADH1

34	104	4.4	550	1	HEMA_IADH2	P12583 influenza a
35	104	4.4	550	1	HEMA_IADH3	P12584 influenza a
36	103.5	4.4	2464	1	MAPB_MOUSE	P14873 mus musculus
37	103	4.4	566	1	HEMA_IADH2	P26135 influenza a
38	103	4.4	1577	1	MYSH_ACACA	P47808 acanthamoeb
39	102.5	4.3	1485	1	TOP2_SCHPO	P08096 schizosacch
40	102.5	4.3	3106	1	IMA2_MOUSE	O60675 mus musculu
41	102	4.3	540	1	YK26_YEAST	P36112 saccharomyc
42	102	4.3	1664	1	INT1_CANAL	P33705 candida alb
43	101.5	4.3	566	1	HEMA_IAME1	P03449 influenza a
44	101.5	4.3	812	1	IF38_YEAST	P32497 saccharomyc
45	101.5	4.3	2748	1	NIM1_YEAST	Q00402 saccharomyc

ALIGNMENTS

RESULT 1	ID	YMH1_YEAST	STANDARD:	PRT:	607 AA.
AC	004632	003641:			
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Hypothetical 69.8 kDa protein in RPL6A-DK1 intergenic region.				
GN	YMO71C.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE OF 1-113 FROM N.A.				
RC	STRAIN=5288C / AB972;				
RA	Bowman S., Churcher C., Barrell B.G., Rajandream M.A.;				
RL	Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 88-607 FROM N.A.				
RC	STRAIN=5288C / AB972;				
RA	Brown D., Bowman S., Barrell B.G., Rajandream M.A.;				
RL	Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	or send an email to license@sib-sib.ch).				
DR	EMBL: Z38114; CA86249.1; -				
DR	EMBL: Z46373; CA86507.1; -				
DR	SGD: S0004536; YMO71C.				
KW	Hypothetical protein.				
SQ	SEQUENCE 607 AA; 69828 MW; 57E34E2D3664B6 CRC64;				
Query Match	5.4%	Score 127;	DB 1;	Length 607;	
Best local Similarity	22.1%	Pred. No. 0.28;			
Matches	90;	Conservative 63;	Mismatches 147;	Indels 108;	Gaps 21;
QY	108	KESDNPANSTKRRFWRRGKNHSYFDLVFNLEKNVTRDADATDI-----EN 158			
DB	137	ENDNDGRIMTTSN---NIAKKKDEDFKALSRLNRKSTKEDDKDRLSRTLVYLEN 193			
QY	159	FAS-----RYLYMATLYYKT-----YTVNDEFGASFNNKLSFT 191			
DB	194	LDSTIDLMELPLARTCIKRTGYOEAVMLYTHTSLRSHPSTIVDEYCEKVLNLS-T 252			
QY	192	TGLFGW-----GIRKALKOIIISNPLDVGTEHSY-----SRLOHTTSYNDY 234			
DB	233	TMLSLVLLKLTNVSNSIKTL-OYLSNIPFDKTKNSLSVFLAMKYKRTIDEIASY 311			
QY	235	MDTOIPALPKFAKRFSL-NVVOQLATVAGYVDTPWYKKWMLKMFV-VNRFIPT 289			

Db 312 PLDVE-----SSNESLIEAMVKKRIEVLREHV-----YMSLNFLEKSLYDTNDLEIP- 359
QY 290 KFFNKREIRPSKALKEKYSTDTKDLFENKIGQGVDFENKRIIDPSKALKEKVSNDAKD 349
Db 360 ---FPEELE---STVLRLNGTNEKEIEKE-----KETKKEEYQKDSVANNEED 404
QY 350 LFEKNIIGGTVDFINNEIRDPKALIRKVSSTGAEDLFENKIGQGVDFINNEIRDPKAL 409
Db 405 VTEKN---SIEDVOEEV---OGKVEGEDGAEKRTETENETENETV---NKTEDKAE-- 450
QY 410 IRKYVTEADDLFENKIGQGVDFINKEI--RDPKALIRKSTENDL 455
Db 451 --KEKEEVENTKDKNAKEKEEIRNKVEYTPPEPSKINSKAKKEEET 496

RESULT 2
FRPC_NEIMC
ID FRPC_NEIMC STANDARD; PRT: 1829 AA.
AC P5127;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated protein frpc.
GN FRPC.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FAM20 / SEROGROUP C;
RX MEDLINE=94018616; PubMed=8412674;
RA Thompson S.A., Wang L.L., Spelling P.F.;
RT "Cloning and nucleotide sequence of frpc, a second gene from
cytotoxins";
RT Neisseria meningitidis encoding a protein similar to RTX
Mol. Microbiol. 9:85-96(1993).
CC -1- FUNCTION: MAY PARTICIPATE IN THE PATHOGENESIS OF MENINGOCOCCAL
DISEASE.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED AND SECRETED.
CC -1- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE RTX PROKARYOTIC TOXIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; L06299; AAA99902.1; -
DR Interpro: IPR001343; Hemlyns_Ca_bind.
DR Pfam: PF00353; hemolysins; 9.
DR PRINTS; PR00313; CABDNNGRPT.
DR PROSITE; PS00330; HEMOLYSIN-CALCIUM; 17.
KW Toxin; Calcium; Outer membrane; Repeat.
FT DOMAIN 879 1702 43 X REPEATS, GLY-RICH.
FT REPEAT 879 884 1.
FT REPEAT 888 893 2.
FT REPEAT 897 902 3.
FT REPEAT 1016 1021 4.
FT REPEAT 1025 1030 5.
FT REPEAT 1034 1039 6.
FT REPEAT 1043 1048 7.
FT REPEAT 1052 1057 8.
FT REPEAT 1061 1066 9.
FT REPEAT 1070 1075 10.
FT REPEAT 1079 1084 11.
FT REPEAT 1088 1093 12.
FT REPEAT 1097 1102 13.
FT REPEAT 1216 1221 14.

FT REPEAT 1225 1230 15.
FT REPEAT 1234 1239 16.
FT REPEAT 1243 1248 17.
FT REPEAT 1252 1257 18.
FT REPEAT 1261 1266 19.
FT REPEAT 1270 1275 20.
FT REPEAT 1279 1284 21.
FT REPEAT 1288 1293 22.
FT REPEAT 1297 1302 23.
FT REPEAT 1306 1311 24.
FT REPEAT 1315 1320 25.
FT REPEAT 1324 1329 26.
FT REPEAT 1333 1338 27.
FT REPEAT 1342 1347 28.
FT REPEAT 1351 1356 29.
FT REPEAT 1360 1365 30.
FT REPEAT 1369 1374 31.
FT REPEAT 1378 1383 32.
FT REPEAT 1387 1392 33.
FT REPEAT 1396 1401 34.
FT REPEAT 1405 1410 35.
FT REPEAT 1414 1419 36.
FT REPEAT 1423 1428 37.
FT REPEAT 1432 1437 38.
FT REPEAT 1441 1446 39.
FT REPEAT 1450 1455 40.
FT REPEAT 1459 1464 41.
FT REPEAT 1468 1473 42.
FT REPEAT 1477 1482 43.
SO SEQUENCE 1829 AA; 197622 MW; 5C3494020A782DC8 CRC64;

Query Match 5.2%; Score 122; DB 1: Length 1829;
Best local similarity 20.3%; Pred. No. 2.3;
Matches 80; Conservative 63; Mismatches 156; Indels 96; Gaps 18;

QY 94 LPDYOLEAFILRKESANPANSTKRFMRFRGRKHSFPHLYVFLLEKNYTRDA-- 151
Db 35 LGDRYADNAAAIYVGKDTNLNGIN---LWM--RKG-----VENLMDDTYGGKTRL 78
QY 152 ---DATDIENFASRLYATLYKYTVNDEFGASFENKLSFTGGLFGGJKRALQIIR 208
Db 79 EKFDYVALQHF--SQYVDLNNKNGRLPMTSELSIYKAVTY-----HGVSAAIDLVIN 133
QY 209 SNLP-----LDIGTE---HSVSLQHTTSYKDYMDQIPALPK-----FAKRS 250
Db 134 RSLPDMADGYVALGIGIAERIHNEQAVNPNNGSEBRNRQLISALDGFDSFEKHF 193
QY 251 LM--VYORLLATVAGYVDTPIYK--KW-----YMLKPMVNRVPIPTKFFNKELREPSK 302
Db 194 FLQSVMDLTKLGYEYTTIDGQKIGMGNGIINDLYSVKREMTGIFEIYNNNIKQNE 253
QY 303 ALKEKVS-----TDTKDLFENKIGQGVDFENKRIIDPSKA 338
Db 254 AFKNEINSLVMDKMAAGKEFGDDLTQNNLTQAAEIIYNDIVNTSGIKRGK-RIKE 312
QY 339 LKEKVSNDAKDLFENKIGQGVDFINNEIRDPKALIRKVSSTGAEDLFENKIGQGVDFI 398
Db 313 LSEKMKNAASDLADGS-----AEKAKVVEDLQAQAKAEVEN--AKSTAEKA 357
QY 399 INNEIRDPKAL--IRKYVTEADDLFENKIGQGVDF 431
Db 358 AQAAREFFKGLPSFKDLAEKFRDLFPNP--EGWID 390

RESULT 3
HEMA_IAUSS
ID HEMA_IAUSS STANDARD; PRT: 566 AA.
AC P03453;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [contains: Hemagglutinin HA1 chain;
Hemagglutinin HA2 chain].

```

GN HA.
OS Influenza A virus (strain A/USSR/90/77).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11481;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84090410; PubMed=6690716;
RA Concanon P., Cummings I.W., Salser W.A.;
RT "Nucleotide sequence of the influenza virus A/USSR/90/77
RN hemagglutinin gene.";
RL J. Virol. 49:276-278(1984).
RN [2]
RP SEQUENCE OF 1-343 FROM N.A.
RX MEDLINE=84074955; PubMed=6196912;
RA Nakajima S., Nakajima K., Kendal A.P.;
RT "Identification of the binding sites to monoclonal antibodies on
RT A/USSR/90/77 (H1N1) hemagglutinin and their involvement in antigenic
RT drift in H1N1 influenza viruses.";
RL Virology 131:116-127(1983).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; K01330; AAA43206.1; ALT_SEQ.
DR EMBL; K01331; AAA43240.1; -.
DR PIR; A04064; HMYUR.
DR HSSP; P03437; 5HMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein; signal.
FT SIGNAL 1 17
FT CHAIN 18 343
FT CHAIN HEMAGGLUTININ HA1 CHAIN.
FT CAROHD 27 566
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 28 28
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 40 40
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 104 104
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 144 144
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 172 172
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 177 177
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 286 286
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 304 304
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 498 498
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 557 557
FT CAROHD N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLCT 13 13
FT CONFLCT A -> S (IN REF. 2).
FT CONFLCT 180 180
FT CONFLCT S -> K (IN REF. 2).
FT CONFLCT 296 296
FT CONFLCT MISSING (IN REF. 2).
SQ SEQUENCE 566 AA; 63460 MW; 352B39299B7C4B4 CR64;

```

```

Query Match 5.1%; Score 121; DB 1; Length 566;
Best Local Similarity 20.8%; Pred. No. 0.63;
Matches 94; Conservative 68; Mismatches 161; Indels 130; Gaps 23;

```

```

OY 33 DMSDLSAVS--DNF-----AERICGOVPRGNSCASVSAYMRCACOD-----CLTLOS 80
DB 117 ELREOLSSVSSEREFEIPFKEK---SWPK-HNVTGCVTASCSHGKSSRYRLNLWTEKN 172
OY 81 LKYPLEAKYQPLTLPDPYOLEAFLFKESDA-----NPANSTKRFPMWFRFRGKNHS 133

```

```

DB 173 GSYV-----NLSSSYNNKKEEVLVLGVHHPNSIEDQKTIYR---KENA 214
OY 134 YRHDLVFNLEKNVTRDADATDIENFASRLYMATLYTYTNRVDEFGSF---FNKLSF 190
DB 215 YVSVSNNYNNRRPTEPIASRPYRGAGRIINYYWTLLEBDITIFBANGLIAPWAFAL 274
OY 191 TTGLFCMGIKRALKQIIRFNLPDIDGTEHSVSLQHTSS--YKQMDIQIAPLPFAAR 248
DB 275 NRG-RESC-----ITTSNASMDECDTCKQTPQCAINSLPQNIHPPTIGECPRVNS 326
OY 249 FSLMVV-----QRLATVAGYVDPW---YKKWY-----M 275
DB 327 TKLRWVTGLRNIPISIQSRLGALGFIETGWTGMIDGYVGHQNEGSGYAADQKSTQ 386
OY 276 KLNKFNVRV-----FIPTKFNKKEIREPSKALKKEXVDTKD-----L 315
DB 387 NAINGTNNVNSYIEKMTQFTAVGKEFNK-DEKREMNLNKKYDDDFDIWYNAELVL 445
OY 316 FENKIGOGTVDFPNKRIKRPDSKALKREKVSNDAKDLFENKIGOGTVDF---INNEIRDPK 372
DB 446 LEN---ERTLDPHDSVKKAYEKVKSQLNNAKE-----IGNCFEYHKCNE----- 491
OY 373 ALIRKYSTGAEDL---FENKIGOGTVDFINNE 401
DB 492 -CMESYKNCTIYDPRKYSSESKLNREKIDGVKLE 523

```

```

RESULT 4
ID HEMA_I4ZUK STANDARD; PRT; 566 AA.
AC P26141;
DT 01-MAY-1992 (rel. 22, Created)
DT 01-MAY-1992 (rel. 22, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain].
GN HA.
OS Influenza A virus (strain A/Swine/Ukkel/71/84).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11517;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114135; PubMed=1731092;
RA Bean W.J., Schnell M., Katz J., Kawaoka Y., Naeye C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 influenza virus hemagglutinin from human and
RT nonhuman hosts.";
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL; M3775; -. NOT_ANNOTATED_CDS.
DR HSSP; P03437; 1HTM.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PR00329; HEMAGGLUTN12.
DR ProDom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein; signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CHAIN HEMAGGLUTININ HA1 CHAIN.

```

[illegible]

Query Match 5.0%, Score 117, DB 1, Length 1038;

Best Local Similarity 20.2%; Pred. No. 4.9;
Matches 81; Conservative 62; Mismatches 150; Indels 108; Gaps 19;

```

OY 94 LDDPQLEAFLFFRESOANPANSTKRRFMRFRGKNHSHYHDLVFNLEKNVTRDADA 153
DB 35 LDDRAADDAALVGGKADNGLN-----LWM-----VENLMDTVGK---K 75
OY 154 TDENFASRYLYMATLYYKTYT-----NVDGFAEFNKLSEFTTGLFGWGIKRA 202
DB 76 TLEKEDR-----VALQHFQYARLNNQNGRLPNTSEIRSYKAVTDN---GVSSSA 127
OY 203 LKQIIRSNLP-----LDIGTE---HSVSRLOHTISSYDYMDQIPALPK----- 244
DB 128 IDLVINRSILPMDADGYWALGIGIEARHNEQAVNNPNSESDNRKQSLALDKGDFGSE 187
OY 245 FAKRSLMNVQRLATVAG--YVDPFWYK--KW---YKKLNPMNVRVFIPTKKFENKE 296
DB 188 KKHFTFLOSVMADYTKLGVEYTTIGWMKIGGNGIINDLKSVYKREMTGIFELVNRN 247
OY 297 IREPSKALKEKVS-----TDTKDLFENKIGQTVDFENKEI 332
DB 248 IKGHDAEAFNEINSLVHOMKAKGKERGDOLNTQMNMLTQAEIYINDIVDNTSQGLEKCV 307
OY 333 RDPFKALEKVSNDAKDLFENKIGQTVDFINNEIRDPKSKALIRKYSTGAEDLFENKIGQ 392
DB 308 K-AIKELSEKMKNAASDLADGS-----AEKAKQVVEDLAQAKABAYEN--AK 351
OY 393 GTVDFINNEIRDPKSKAL--IRKRYTEADDLFENKIGQTVD 431
DB 352 STAEKAAQAAREFFKGLPSFKDLAEKFRDLFPNP--EGWID 390

```

RESULT 7

```

Y162_UREPA
ID Y162_UREPA STANDARD: PRT: 457 AA.
AC O9POX2:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein UY162.
GN UY162.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3:
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Caswell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
CC DR EMBL; AE002116; AAF30568.1; -
CC InterPro: IPR000087; Collagen.
KW Hypothetical protein; Complete proteome.
SEQUENCE 457 AA: 51904 MW: 57D0ERC8D0A4D637 CRC64:

```

Query Match 4.9%; Score 116; DB 1; Length 457;
Best Local Similarity 19.4%; Pred. No. 1;
Matches 83; Conservative 61; Mismatches 133; Indels 150; Gaps 19;

78 LOSLKYPLEAKQPLTLPDPYOLEAFLFKES---DANPANSTKRRFMRFRGKNHSHY 134

```

DB 25 LEDNQLERKQALT-----TFILRAQSMIDCVASNG-----GAGFIYSM 66
OY 135 FHDIVFNLEKKNVTRADAT-----
DB 67 FDDLKTTDYOKTRFEKALITLTNYLWLGHPIDKDYTAIGDGLSIFKSENADEPKRL 126
OY 155 DIENFASRYLYMATLYYKTYT-----YTVDGFAEFNKLSEFTTGLFGWGIKRA 204
DB 127 DLIIN-----KLVRGGLKSKKPFANFNDDGLIYVVDLKAIAIN-----NVHEELR 171
OY 205 Q-----TIRSNLPIDIGTEHSVSRLOHTISS-----KYQYM--DTQIP--ALPKRA 246
DB 172 DHYISKNNPNNVNNMEDLN---HHNINNVSSINASGLATLNGFTLIDCTLPPDLTTDKSN 228
OY 247 KRPESLMNVQRLATVAGYVDFPFWYK--WYMKLNPMNVRVFIPTKKFENKEIREPSKAL 304
DB 229 KYVD-----NATANYVDIHDKMLAWYGGONNNALTR-----QELYEICNKI 270
OY 305 KEKVSFDTKDLFENKIGQTVDFENK-----EIRDPKSKALKEKVS- 345
DB 271 NSKMYDA--IYDYKAGELVYDLDEKNTIKYKALVNNGVSTHNIYDPSKQINENWEL 328
OY 346 -DAKDLFENKIGQTVDFINNEIRDPKSKALIRKYSTGAEDLFENKIGQTVDFINNEIRD 404
DB 329 INIQIDLNKIRKEGVQDAVDNKNVEYKKAIDTKENTFTKNL-ENKVNKAQAVLSSEYED 387
OY 405 PSKALIR 411
DB 388 KASGVK 394

```

RESULT 8

```

MLP1_YEAST
ID MLP1_YEAST STANDARD: PRT: 1875 AA.
AC 00245:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C:
RX MEDLINE=93247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure."
RL Mol. Gen. Genet. 237:359-369(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94205265; PubMed=8154186;
RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MLP1 genes and three
RT new open reading frames."
RL Yeast 9:1349-1354(1993).
CC -----
CC -I- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
CC REPAIR.
CC -----
CC -I- SIMILARITY: SOME, TO THE TPR ONCOGENE.
CC -I- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MLP1".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).

```


CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDPA ATPASE
CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CIS AT THE SH-1
CC POSITION (686).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M14628: AAA3327.1; -.
DR PIR: A26655; A26655.
DR PIR: S00250; S00250.
DR PDB: 1WMA: 03-DEC-97.
DR PDB: 1AMD: 17-AUG-96.
DR PDB: 1MMD: 03-DEC-97.
DR PDB: 1MMN: 03-DEC-97.
DR PDB: 1MND: 17-AUG-96.
DR PDB: 1MNE: 17-AUG-96.
DR PDB: 1VOM: 23-DEC-96.
DR PDB: 1LVK: 28-JAN-98.
DR D1CVDDB: D001008; mbca.
DR InterPro: IPR000048; IQ.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR Prosite: PS00096; IQ; 1.
DR Myosin: Colled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
FT NP_BIND 817 791 IQ.
FT NP_BIND 817 2116 COILED COIL (POTENTIAL).
FT DOMAIN 179 186 ATP.
FT DOMAIN 638 660 ACTIN-BINDING.
FT DOMAIN 738 752 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678 ALKYLATION (SH-1).
FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770B1E56A1 CRC64;
Query Match 4.8%; Score 113.5; DB 1; Length 2116;
Best Local Similarity 20.0%; Pred. No. 10;
Matches 91; Conservative 77; Mismatches 166; Indels 121; Gaps 23;

QY 261 VAGYDTPYKMKWYKLFNPNRVETPTKPFENKREPSKAL-----REK 307
DB 802 YIDFSKMPWV-KLFSKARLL-----KRNFEKIKKEEILELKNLSDSTQK 853
QY 308 VSTDRKDLFENKIGOGTVDFPNKREIRPSKALKE--KVSNDARDLFENKIGOGTVDFINN 365
DB 854 LEKSLKDTRESNVL-----DLQRLKAEKETLKAYDSKDALE----- 890
QY 366 EIRDPKALIRKVSIGAEEDLFENKIGOGTVDFINNIRDPKALIRKYTTEADDFENKI 425
DB 891 -----AKRELEIRVEQM--ESELDEKLALEN--LQNKRSVEKRVNDELEDEEOK 940
QY 426 GOGTVDFINKEIRDPKALIRKVSIGAEEDLV--LEK 458
DB 941 LNTLEKLLKKYEELEEMKRVNDGSDPTISRLK 975
RESULT 10
MAPB_RAT STANDARD; PRT; 2459 AA.
ID MAPB_RAT
AC P15205: 062958: Q9ER21: Q9QW92:
DT 01-APR-1990 (rel. 14, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RA MEDLINE=96257242; PubMed=8666295;
RX Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain; and Gliat tumor;
RA MEDLINE=92347374; PubMed=1639092;
RX Zauner W., Kratz J., Staunton J., Felck P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RA MEDLINE=90059871; PubMed=2555150;
RX Rlenitz A., Grentingloh G., Hermans-Borgmeyer I., Kirsch J.,
Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
nervous system that is immunologically related to microtubule-
associated protein 5.";
RL Embo J. 8:2879-2888(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
in rat CNS and PNS during development.";
RL J. Neurosci. Res. 49:319-332(1997).
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.

FT CHAIN 346 566 HEMAGGLUTININ HAZ CHAIN.
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 566 AA: 63514 MW: AEC60738C765F637 CRC64;

Query Match 4.7%; Score 112; DB 1; Length 566;
 Best Local Similarity 20.5%; Pred. No. 2.5;

Matches 93; Conservative 68; Mismatches 162; Indels 130; Gaps 23;

QY 33 DMSDYLAVS--DNF---AERICSOVPGKSGNSASVYMSRCADKOD-----CITLOS 80
 DB 117 ELREOLSSVSFEPREIPFKER---SWPK-HNVTGCVTASCCHKGKSPFYRNILMTKRN 172
 QY 81 LKYPLEAKYQPLTLPDPYQLEAFLFKESDA-----NPANSTPKRRMRRRGRKNS 133
 DB 173 GSYF-----NLKSYVNNKKEKVLVMGVIHPSNIEDQKTYR---KENA 214
 QY 134 YPHDVLNLEKKNTRDADATDIENFASRYLYMATLYKTYTNVDEFGASF--FNKLSF 190
 DB 215 YSVSVSNVNRFPTEPEIARPKVROGAGRINYYWTLLEGGDITIFANCNLLAPMAFL 274
 QY 191 TGGTGGWGKRAKLRKQIIRNLEPLDIGTEHSVRLQHTSS--YKDYMDIQIPALPKPAK 248
 DB 275 NRG-EGSG-----ITNSMASDECDTRQCPGAINSLSPFONIHPTIGCEPKYVS 326
 QY 249 FSLMVV-----OBLATVAGYVDTPW---YKKVY-----M 275
 DB 327 TILRVATGLRNIPSTOSRGLFGAIAGFIGGWTGMIDGVIYNHNGEAGSVAADOKSQ 386
 QY 276 KUKNFMVNRV-----FIPYKRFNKETREPSKALKKESVDTKD-----L 315
 DB 387 NAINGITNKNVSVIEKMTQFQAVGKEFD-ELEKRMENLNKKVDOGFLDIWYNAELVL 445
 QY 316 FENKIGGQVDFENKRIKPSKALKKESVNDAKDLFENKIGGQVDF--INNEIDPSK 372
 DB 446 LEN---ERLIDPHDSNVKMLYEKVSQLEKNNAKE-----IGNGCEFFYHKCNNE----- 491
 QY 373 ALIRKYSTGAEDL---FENKIGGQVDFINNE 401
 DB 492 -CMESYKNGTYDYPKYSKSLNREKIDGKLE 523

RESULT 12
 YD86_SCHPO STANDARD: PRT: 1957 AA.

AC 010411:
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-OCT-1996 (rel. 34, Last annotation update)
 DE Hypothetical 222.8 kDa protein Clp3.06C in chromosome I.
 GN SPAC13.06C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----

DR EMBL: 270690; CAA94624.1;
 KW Hypothetical protein.

SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;

Query Match 4.7%; Score 112; DB 1; Length 1957;
 Best Local Similarity 19.2%; Pred. No. 11;

Matches 107; Conservative 94; Mismatches 204; Indels 152; Gaps 27;

QY 9 DWTKTLAASEVDSAAANY-MINSDSDYLSAVSDNFA-----BRICSOVPGKSN 58
 DB 205 DLSROLTYTEKIDKKEXYERIKEDVSSIKMSLAEGQSNKSLRGEQRL-EKLLVSN 263
 QY 59 CSASVSAVMSRCAKQDCLTLQ-----SLKYPLEAKYQPLTLPDPYQLE 101
 DB 264 KTVSTLRQTENSIRACBCKTLQEKLEKCAINEDSKLLEELKHNV-ANYSDAIVHKDLIE 322
 QY 102 AAFLEKESDANPANTERKFRFRRC-----NHSYFDLVNMLE 144
 DB 323 DISTRISEFD---NLKSRDITLSTINEKLEKLRNTIGSLKDSRTSNQLEEMVELKE 378
 QY 145 KNVTRPADATDIENFASRYLYMATLYKTYTNVDEFGASFNKLSPFTGLFGWGIKRAK 204
 DB 379 SNRTINSQLTDAESKLSF---EQEKSLKGSIDEVQ---NLSKSKDM---VKQVSS 427
 QY 205 QI-IRSNLEPLDIGTEHSVRLQHTSSYKDYMDIQIPALPKPAK--FSLVAVOURLAT 260
 DB 428 QLEAARSSL-----AHAAGKLAELINSE-RDFQNNKIKPFKEIEDLRACLNSMSNELKE 480
 QY 261 VAGVYDTPWYKWKYMLKPMVNRVFIPTKRFNKETREPSKALKKESVDTKD--OLEFN 318
 DB 481 KSLALD-----KKDOELNLN---REQIEQKVSSTQSSLSQSLQDLITNEKKKHEVYS 532
 QY 319 KI-----GGQVDFENKE-----IRPSKALKKESVSN 345
 DB 533 QINELKGLQTELSNEHLSQSLTSLAEKAAVATNNELSSKNSLQTLCAVFOKRLAK 592
 QY 346 DAKDLFEN-----KIGQGVDFINN-----EIRPSKALKKYSTGAEDLFE 387
 DB 593 SYWQLEKNQONSLSDTSPFKLINESHOELNNHQTITKOLKPTSSKLOLOLERAN--FE 650
 QY 388 NIGQGVDFINNEIR-----DPSKALIRKYVTEADDLFENKIGQGVDFINKEIRDP 440
 DB 651 QR-ESTLSDENNDLTKLLKLEESNKSILKR-QEDVDLERN-----TQTLKEDLRKS 701
 QY 441 SKALIRKYSTEADNMLE 457
 DB 702 EEARL-RFSKLEAKNLRE 717

RESULT 13
 ALP6_SCHPO STANDARD: PRT: 832 AA.

AC 09USQ2: 09P954;
 DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Spindle pole body component Alp6.
 GN ALP6 OR SPBC902.01C OR SPBC428.20C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A. AND CHARACTERIZATION.
 RA MEDLINE=20532503; PubMed=11080156;
 RX Vardy L., Toda T.;

RT "The fission yeast gamma-tubulin complex is required in G(1) phase and
 RT is a component of the spindle assembly checkpoint."
 RT EMBL J. 19:6098-6111(2000).
 RN [2]
 RP SEQUENCE OF 1-566 FROM N.A.
 RC STRAIN-972;
 RA Saunders D., Harris D., McDougall R.C., Rajandream M.A., Barrell B.G.,
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF THE GAMMA TUBULE COMPLEX THAT IS REQUIRED
 CC FOR THE REGULATION OF BOTH INTERPHASE MICROTUBULES AND MITOTIC
 CC BIPOLAR SPINDLES.
 CC -1- SUBCELLULAR LOCATION: SPINDLE POLE BODY AND THE MICROTUBULE
 CC ORGANIZING CENTER (MTOC).
 CC -1- SIMILARITY: BELONGS TO THE SPB ALP6/SPC98 FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
 CC FRAMESHIFT IN POSITION 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB040811; BAA94097.1; ALT_FRAME.
 CC EMBL: AL133306; CAB62095.1; -.
 CC DR Microtubules; Mitosis.
 CC KW SEQUENCE 832 AA; 95996 MW; C92771C3DF5C01A CRC64;
 SQ

Query Match 4.7%; Score 111.5; DB 1; Length 832;
 Best Local Similarity 20.0%; Pred. No. 4.3;
 Matches 101; Conservative 80; Mismatches 194; Indels 129; Gaps 23;

QY 4 SDSVDYTTTLAASEVDSANANAYMNSDYLAVSDNFAERICSOVPGK----- 56
 DB 107 SSSSRVSSHLT--DEGISPIPISTEEV-----SSNFQOTRYDQVDPENDQITDMD 156
 QY 57 ---SNCSASVASYMRCAKQ-----DCLTQSLKPLAKQPLPLPPYQL 100
 DB 157 EGIENESSISIAHDSRLNKTSTSSVQHTLITLADLLSISYVLOG-----1 204
 QY 101 EAAFIIFKESDANPANSTKRFMMRRFR--GKNHSYFHDL-VFNILKKNYTRDADATDIE- 157
 DB 205 STREYVQFKNEIALSKRIPQVYLLQMRALSETGLYQELKVFENYDPSVSQSIDGNVSK 264
 QY 158 ---NFASTRILY-MATLYITYTNVDEFGASFENKLSFTTGL-----FGWGIKRAL 203
 DB 265 AFINDOSLALQSLKSVISKELTNFLALIASLDSQIRADASLEKPMYTIKRCIMATOVAKL 324
 QY 204 KOIIRSNLPLDIGTEHSVSLQHTSSYKDYMDTOIPALPKFAKRRSLMWVQRLATVAG 263
 DB 325 KLRILSSVANDNMQENKRLIOVSKYVHGDPILIOELIS-----DKILTEITG 373
 QY 264 YVDTPWY---KKWYMLKKNFMNVRVEIPTKKFENKEIRPSKALKKEKYSTDTKDFENKI 320
 DB 374 ----PLYEMIEHWYK-----GELVDPYQGEF-----VKKKNSESHD-----HQ 409
 QY 321 GGGTV-----DFNKEIRDP--SKALKKYSNDAKOLFENKIGGVVDPIINNEIRDSKA 373
 DB 410 GGGDVVWKRYFLDKELIPSLSEELVYDKIFLIGKSLNFARVCGCDFD-----AOEYOK 465
 QY 374 LIRKVSAGADELFENKIGGVVDPIINNEIRDSKALIRKYTEADDFENK-----IGGGT 429
 DB 466 LVKKLSYRDPHSLLEYVDKAYTESINHLV-----YLMEEVFLTHDKALKATKYLLOGG- 519
 QY 430 VDFIN-----KEIRDPKALIR 446
 DB 520 -DEVDLMEISLQSLDQNPANTLFR 542

RESULT 14
 ORC1_KL01A

ID ORC1_KL01A STANDARD; PRT; 886 AA.
 AC P54788;
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 01-NOV-1997 (rel. 35, Last annotation update)
 DE Origin recognition complex subunit 1.
 GN ORC1.
 OS Kluyveromyces fragilis (yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96099401; PubMed=7502077;
 RA Gavin K.A., Hidaka M., Stillman B.;
 RT "Conserved initiator proteins in eukaryotes."
 RL Science 270:1667-1671(1995).
 CC -1- FUNCTION: COMPONENT OF THE ORIGIN RECOGNITION COMPLEX (ORC) THAT
 CC BINDS ORIGINS OF REPLICATION. IT HAS A ROLE IN BOTH CHROMOSOMAL
 CC REPLICATION AND MATING TYPE TRANSCRIPTIONAL SILENCING. BINDS TO
 CC THE ARS CONSENSUS SEQUENCE (ACS) OF ORIGINS OF REPLICATION IN AN
 CC ATP-DEPENDENT MANNER.
 CC -1- SUBUNIT: ORC IS COMPOSED OF SIX SUBUNITS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE ORC1/CDC6/CDC18 FAMILY.
 CC -1- SIMILARITY: SOME TO YEAST SIR3.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U40151; AAC49130.1; -.
 CC DR InterPro: IPR003959; AAA_subfam.
 CC DR InterPro: IPR001025; BAH.
 CC DR Pfam: PR01426; BAH; 1.
 CC DR Pfam: PR01426; BAH; 1.
 CC DR SMART: SM00439; BAH; 1.
 CC KW DNA replication; Nuclear protein; DNA-binding; ATP-binding.
 CC FT NP_BIND 471 478 ATP (POTENTIAL).
 CC SQ SEQUENCE 886 AA; 101393 MW; 2DA65EC9291F182F CRC64;
 SQ

Query Match 4.7%; Score 110.5; DB 1; Length 886;
 Best Local Similarity 19.5%; Pred. No. 5.4;
 Matches 99; Conservative 82; Mismatches 198; Indels 129; Gaps 23;

QY 5 SDSVDYTTTLAASEVDSANANAYMNSD--MSDYLSAVSDNFAERICSOV----- 53
 DB 274 EALSDNESDSEYHESKEEPPANSSSDSEDFEDYQSAEELAYVERAKKVRIRIKDIP 333
 QY 54 -PRGNSCSASVASYMS---RCANQDCLTLQSLKRYPLEAKYQ-PLTLPDPYQLBAFLFK 108
 DB 334 SPVKSQTPLOPSAVHSSPRKFNINIVRAKAVTSPSKRYKNPKDIPDLND-----IFQR 388
 QY 109 ESPDANPANSTKRFMMRRFRGKNHSYFHDVLFULEKNYTRD-ADATDIEFN-ASRIYIM 166
 DB 389 HNNDDLIALEERFRFVSAGKMETTFISKVKKOLNSNSKEELVAKADFNQYLPARENFE 448
 QY 167 ATLTYTNTVDEFGASFENKLSFTTGLFGWGIKRALKOIIRSNLPLDIGT---EHSVSR 223
 DB 449 ASIYLSLSAI-EAGTISTSIYAGTGV---GTLIVREYVK-----DLMTSADQELR 499
 QY 224 LOHI-----TSYK-----DYMDTOIPALPKFAKRRSLMWVQRLATVAG 266
 DB 500 FOYIEINGLKIYKASDSYEVFMOKISGEKLTSGAAMESELEFYFNKYVPATKKRIYVLDE 559
 QY 257 LLATVAGYVDTPW-YKKW--YMKLKNFMV---NRVFIPTKKFENK----- 295
 DB 560 LDALVSKDSQDVVYVNFPMATYVSNAKLIVAAVANTLDPERHNLKNTISSRIGFTIRIMETGY 619

QY 296 -----EIR-----EPKSKALEKSTDTKLEFNKIGOGTVDFENKIRDP 335
DB 620 THEELRTIINLRKYLNESEFVDEPCTSSWYISPDSSSTI-ETDEEKRPDPSN----- 672
QY 336 SKALEKESYNKADKLEFNKIGOGTVDFINNEIRDPKSLIRKVSIGADL-----FENKI 390
DB 673 YKRLKLRINPAIEIASKRIAS-----VSGDVRALKVYKRAVEYENDYLRKLYERLY 727
QY 391 -----GOGTVDFINNEIRDPKSL 409
DB 728 NSKKTSGNGNGNELOSVEIKHITKAL 755
RESULT 15
BACC_BACLI STANDARD: PRT: 6359 AA.
ID BACC_BACLI
AC 068008:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Bacitracin synthetase 3 (BA3) [Includes: ATP-dependent isoleucine
adenylase (IleA) (isoleucine activase); ATP-dependent D-phenylalanine
adenylase (D-PheA) (D-phenylalanine activase); ATP-dependent histidine
adenylase (HisA) (histidine activase); ATP-dependent D-aspartate
adenylase (AsnA) (aspartate activase); ATP-dependent asparagine
adenylase (AsnA) (asparagine activase); Aspartate racemase
(EC 5.1.1.13); Phenylalanine racemase [ATP hydrolysing]
(EC 5.1.1.11)].
DE BACC
GN Bacillus licheniformis.
OS Bacillus licheniformis; Bacillus/Clostridium group;
OC Bacteria; Firmicutes; Bacillus/Staphylococcus group; Bacillus.
CC Bacillus/Staphylococcus group; Bacillus.
NX NCBI_TaxID=1402;
RN 111
RP SEQUENCE FROM N.A.
RX STRAIN-ATCC 10716;
RX MEDLINE-96089193; PubMed-9427658;
RA Konz D., Klenz A., Schoerendorfer K., Marahel M.A.;
RT "The bacitracin biosynthetic operon of Bacillus licheniformis ATCC
10716: molecular characterization of three multi-modular peptide
synthetases";
RT Chem. Biol. 4:927-937(1997).
PL Chem. Biol. 4:927-937(1997).
CC -1- FUNCTION: INDUCES PEPTIDE SYNTHESIS, ACTIVATES AND INCORPORATES
FIVE AMINO ACIDS, FORMS A THIAZOLIDINE RING BETWEEN THE FIRST TWO
CC AMINO ACIDS AND INCORPORATES A D-GLUTAMINE IN THE FOURTH POSITION.
CC -1- CATALYTIC ACTIVITY: L-aspartate -> D-aspartate.
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine -> AMP + diphosphate + D-
phenylalanine
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTHETHEINES
(POTENTIAL).
CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
ANTIBIOTIC BACITRACIN.
CC -1- SUBUNIT: LARGE MULTISUBUNIT COMPLEX OF BAI, BAI2 AND BAI3.
CC -1- DOMAIN: CONSISTS OF FIVE MODULES WITH TWO EPIMERIZATION DOMAINS IN
THE SECOND AND FOURTH MODULES, AND A PUTATIVE C-TERMINAL
THIOESTERASE DOMAIN. EACH MODULE INCORPORATES ONE AMINO ACID INTO
THE PEPTIDE PRODUCT AND CAN BE FURTHER SUBDIVIDED INTO DOMAINS
RESPONSIBLE FOR SUBSTRATE ADENYLATION, THIOALATION, CONDENSATION
(NOT FOR THE INITIATION MODULE), AND EPIMERIZATION (OPTIONAL), AND
N METHYLATION (OPTIONAL).
CC -1- MISCELLANEOUS: BACITRACIN IS A MIXTURE OF AT LEAST TEN CYCLIC
DODECAPETIDES, THAT DIFFER BY ONE OR TWO AMINO ACIDS. THE MOST
ABUNDANT IS BACITRACIN A, A BRANCHED CYCLIC DODECAPETIDE. IT
CONTAINS AN N-TERMINAL LINEAR PENTAPEPTIDE MOIETY (ILE-CYS-LEU-D-
GLU-ILE) WITH AN ISOLEUCINE-CYSTEINE THIAALOID CONDENSATION
PRODUCT AND A C-TERMINAL HEPTAPEPTIDE RING (LYS-D-ORN-ILE-D-PHE-
HIS-D-ASP-ASN), IN WHICH THE FREE ALPHA-CARBOXY GROUP OF THE C-
TERMINAL ASN IS BOUND TO THE EPSILON-AMINO GROUP OF LYSINE. IT
CONTAINS FOUR AMINO ACIDS IN THE D-CONFIGURATION (GLU-4, ORN-7,
PHE-9, AND ASP-11).
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
FAMILY.
CC -1- SIMILARITY: CONTAINS 5 ACTYL CARRIER DOMAINS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF007865; AAC06348.1; -
DR HSSP: P14687; 1AMU.
DR InterPro: IPR000873; AMP-bind.
DR InterPro: IPR001242; DGF4.
DR InterPro: IPR003880; Phosphopant-attach.
DR InterPro: IPR001031; Thioesterase.
DR Pfam: PF00501; AMP-binding; 5.
DR Pfam: PF00668; Condensation; 7.
DR Pfam: PF00550; pp-binding; 5.
DR Pfam: PF00975; Thioesterase; 1.
DR PRINTS: PR00154; AMPBINDING.
DR PROSITE: PS00012; PHOSPHOPANTHETHEINE; 4.
DR PROSITE: PS00455; AMP-BINDING; 5.
DR PROSITE: PS50075; ACP_DOMAIN; 5.
DR Ligase; isomerase; Hydrolyase; Antibiotic biosynthesis;
KW Phosphopantetheine; Multifunctional enzyme; Repeat.
FT REPEAT 461 1034 DOMAIN 1 (ISOLEUCINE-ACTIVATING).
FT REPEAT 1517 2064 DOMAIN 2 (D-PHENYLALANINE-ACTIVATING).
FT REPEAT 2999 3570 DOMAIN 3 (HISTIDINE-ACTIVATING).
FT REPEAT 4047 4612 DOMAIN 4 (D-ASPARTIC ACID-ACTIVATING).
FT REPEAT 5549 6129 DOMAIN 5 (ASPARAGINE-ACTIVATING).
FT DOMAIN 1 966 1033 ACYL CARRIER (ACP) 1.
FT DOMAIN 2 1998 2064 ACYL CARRIER (ACP) 2.
FT DOMAIN 3 3502 3569 ACYL CARRIER (ACP) 3.
FT DOMAIN 4 4544 4610 ACYL CARRIER (ACP) 4.
FT DOMAIN 5 6052 6119 ACYL CARRIER (ACP) 5.
FT BINDING 996 996 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 2028 2028 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 3532 3532 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 4574 4574 PHOSPHOPANTHETHEINE (BY SIMILARITY).
FT BINDING 6082 6082 PHOSPHOPANTHETHEINE (BY SIMILARITY).
SQ SEQUENCE 6359 AA: 722923 MW: 8242736546253074 CR64;
Query Match 4.7%; Score 110.5; DB 1; Length 6359;
Best local similarity 20.5%; Pred. No. 61;
Matches 117; Conservative 58; Mismatches 173; Indels 223; Gaps 30;
QY 7 VGDVTKTLTAASESVDASANAAMNINSDMSDYLSA-----VSDNFAERICSOVPRKSN 58
DB 1897 LGELINQLLKL-DKIDEAAVYARKDDHSDYLCAIYSKEDMTSTSEMLEKELP----- 1951
QY 59 CSASVAVMSRCARODCTTLOSLKYPLEAKYQPLTPDPYOLEAFLFKESDANPANST 118
DB 1952 -HYMPAFYVR-----LKLPLTSDNDYDKRALPAPBDHVAIVGAYEA-----PRNDT 1998
QY 119 EKRFEMRRR-----GKNHSYF-----HOLVPLEKN 146
DB 1999 EAKLVDIRVDYAGAGDISHHFFAAGDSIKALQIYSLRLGKLEMKLDFANPRKD 2058
QY 147 -----VTRDADTDIENFASRYLYUATLYKYTVNDEGASF----- 184
DB 2059 LAKYVKOSORKNANTITVGAHELPIQKW-----YFA-----NKKELDHNOGFVLEFK 2109
QY 185 -----FNKL-----SPTTGLFGW-----GIRALK 204
DB 2110 GGFDESCVKAAPFKIMEGHDLRMVIEKGGDFQYNSNPFEDLFDLVYDVRGDRQAE 2169
QY 205 QI-----IRSLPLDIGT-----EHSVSRQHTSS-----KYQYMDTQI 239
DB 2170 KYVELATSIQKLSSIRKSKLVHLGIFRADEGDHLIIVHLLVVDGVSRLIEFDEFELYS 2229
QY 240 PAL-----PKFAKRFSLMVVORLAVTAVAGYVDTPWYKKYKMYKLNPMVRY 285
DB 2230 QALKGQTLIEIGYKTDSDIOEFARLRLKAYAHSRITLSKEAY-----W-----RINAKRV 2277

QY 286 -FIPFKFKEIREPSKALKEKYSTD--TKDLFENKIGGTVDFENKEIRDPKALKEKY 343
DB 2278 RFLPPNVLKEDYENSTLSTLKKEATDRLRN-----TNRAYTEIND--ILLTAL 2329
QY 344 SDAKDLE-ENRI-----GGGVDFINN-EI-----RDPKALIRKYST 380
DB 2330 LGAGRTIGENLKVMMEHGREDILEGVDITRTIGFTTMYVLADAGEKALSSQIKM 2389
QY 381 GAEDL--FENK-IGGTV-----DFINNE 401
DB 2390 VKETLKRIPNKGIGYGLTKYMAEDDPFTNDE 2420

RESULT 16
HEKA_IAPUE STANDARD; PRT; 566 AA.
AC P03452; Q83964;
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
HA.
GN Influenza A virus (strain A/Puerto Rico/8/34).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11455;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82013600; PubMed=7278968;
RA Winter G., Fields S., Brownlee G.G.;
RT "Nucleotide sequence of the haemagglutinin gene of a human influenza
RT virus H1 subtype.";
RL Nature 292:72-75(1981).
RN [2]
RP SEQUENCE OF 18-343 FROM N.A.
RC STRAIN-SUBTYPE H1;
RX MEDLINE=83129356; PubMed=6186384;
RA Caton A.J., Brownlee G.G., Yewdell J.W., Gerhard W.;
RT "The antigenic structure of the Influenza virus A/PR/8/34
RT cell 31:417-427(1982)."
RL
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: V01088; CAA24272.1; -
CC EMBL: J02144; AAA43194.1; -
CC PIR: A04063; HMTV.
CC HSSP: P03437; 5HMG.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR PRINTS: PR00329; HEMAGGLUTN1.
DR ProDom: PD000225; Hemagglutn; 1.
KM Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 17
FT CHAIN 18 343 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 345 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 146 147 A -> E (IN SUBTYPE H1).
FT VARIANT 156 156 S -> P (IN SUBTYPE H1).
FT VARIANT 200 200 D -> E (IN SUBTYPE H1).
FT VARIANT 204 204 I -> L (IN SUBTYPE H1).
FT VARIANT 208 208 R -> M (IN SUBTYPE H1).
FT VARIANT 269 269 F -> Y (IN SUBTYPE H1).
FT VARIANT 309 309
SQ SEQUENCE 566 AA; 63381 MW; E7AE80403547FCE9 CRC64;

Query Match 4.7%; Score 110; DB 1; Length 566;
Best Local Similarity 20.8%; Pred. No. 3.3;
Matches 97; Conservative 67; Mismatches 171; Indels 132; Gaps 24;

QY 23 SAANAVMINSMSDY-----LSAVSDNFAERISQVPRGS-----NCSASVAYSRCRA 71
DB 101 NSENGICYPEDFDIYELRQOLSSVSS--FER-ELFPKSSPNNTTGTVAACSHAG 157
QY 72 KOD-----CLTQSLKYPLEAKYQPLTLPDYQLEAFLIFRESDA-----NPANST 118
DB 158 KSEFYRNLLMLTEKESYP-----KLKNSYNNKKGKGYLVLMGIHPSNSK 203
QY 119 EKRFMMFRGRKNSYFHDLVFNLEKNYTRDADDIENFASRYLYMATLYKTTTND 178
DB 204 DQD---NIYQEN-AYVSVTSNYSNRRFPELIERPVROQARMYWLRLKPGDTIF 259
QY 179 EFGASFF-NKLSPTTGL-FCWGIKRALKOTIRSNLPLDICTEHSVSLQHTSS--YKDY 234
DB 260 EAAGNLAPRYATAPALSRGSG-----ITTSNASHCEHTCGQPLGAINSLPQNT 312
QY 235 MDQIIPALPKFAKRFSLMVV-----QRLATVAGYDTPW--YKKVY----- 274
DB 313 HPVTIGECPRYVSAKRLMYVTLGRNIPSIQSRDLFGAIFIGCWGMIDGWYGHQNN 372
QY 275 -----MKLNPNVNV-----FIPFKFKEIREPSKALKEKYSTD 311
DB 373 EGGSGYAADQKSTQMAINGITNKNVSYIEKMNIQFNAVGEFKN-LEKREINLKKRYVDG 431
QY 312 TKD-----LFENKIGGTVDFENKEIRDPKALKEKYSTDAKDLEFNKIGGTVD 361
DB 432 FLDIYTNMELLVLEN---ERTLDHDSNVKMLYEYKQLNNAKE-----IGNGCFE 483
QY 362 F---INNEIRDPKALIRKYSTGAEDL---FENKIGGTVDFINNE 401
DB 484 FYKCKNE-----CMESVRNCTYDYPKYSSESKLNREKVDGKLE 523

RESULT 17
THRT1_SCHPO STANDARD; PRT; 577 AA.
AC 009684;
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE Nuclear fusion protein tht1.
GN THRT1 OR SPAC13C5.03.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98106170; PubMed=9442101;
RA Tange Y., Horio T., Shimanuki M., Ding D.O., Hiraoka Y., Niwa O.;
RT "A novel fission yeast gene, tht1+, is required for the fusion of
RT nuclear envelopes during karyogamy.";
RL J. Cell Biol. 140:247-258(1998).
RN [2]


```

Db 689 FNNNSDSSAQTFLTKTAATAITSDNEPTREKNNQTKLTTEVEKAKELVEKAKATY 748
Qy 227 TTSXKDYDIOIPALPKAKRPSLAVVORL-----ATAGYDTPWKYKTKLKN 279
Db 749 SSSQYSEITLLKSSQL-----NDLDDLIJLLTDSGIRGTAVANI-----FKMWYKKN 799
Qy 280 FVNRVFIPTKKE-----FNKEIRP-----SKALKEKYSTDKDF 316
Db 800 STNNEDTSNKEKEGFSFNDLVKQALYIRSMQNLTSKREGRYTDLSVNSTMLQW 859
Qy 317 ENK-----IGQGVDF-----FNKEIRDP-----KALKEKVSND-----RD 349
Db 860 OKKAAQSHTSVNTQTLDDLAKKAFKELEDPNQDAEYKMYRQLQMLMYKGAQNYKN 919
Qy 350 LEENKIGQGVDFINNET--RDPKALIRKYSTGAEJLEPKIG-----QGT-VD 396
Db 920 LLOQALPICTRAFVSWTVGYDKNPSATVSOQKTSSTSSANENPFNPLQNPNTQGSSEIN 979
Qy 397 FINNE---IRDPKALIRKYTEADDFENKI-----GGQTV 431
Db 980 WENDKOTPIQDPSLSENTYRFTDEPFNNNSVALSKNSGSSD 1022

RESULT 19
HEMA_IKIE
ID HEMA_IKIE STANDARD: PRT: 566 AA.
AC P18875:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAZ chain].
GN HA.
OS Influenza A virus (strain A/Kiev/59/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
NCBI_TaxID=11422;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86186951; PubMed=3964310;
RA Beklemishev A.B., Bilnov V.M., Vasilenko S.K., Golovin S.Y.,
RA Karginov V.A., Mamayev L.V.;
RT "Primary structure of the full-size DNA copy of the hemagglutinin
RT gene of Influenza virus A/Kiev/59/79 (H1N1).";
RL Bioorg. Khim. 12:375-381(1986).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M38353; AAA3172.1; -.
DR HSRP: P03437; 2HMC.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KM Envelope protein; Hemagglutinin; Glycoprotein; signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HAZ CHAIN.
FT CAROHD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```

FT CAROHD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 177 177 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 286 286 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 498 498 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63606 MW; A78A51A4F84D9E74 CRC64;

Query Match
Best local similarity 20.5%; Score 108; DB 1; Length 566;
Matches 95; Conservative 66; Mismatches 151; Indels 152; Gaps 23;

Qy 33 DMSDYLSAVS--DNF---AERICQVPGKSNCSASVAYMSKAD-----CLTLOS 80
Db 117 ELREQLSVSSFEFRFIPKER---SWPK-HNVTREGVASCSSHKGSFYNLLMLEKN 172
Qy 81 LKYPLEAK-----YQPLPLDP-----YQLEAFIILFKEDANPANTE-- 119
Db 173 GSTPNLSKSYVNKKREVLVLWGVIHSPNIEDQTYRKRENAVSVSSNRYRFPPELA 232
Qy 120 KRFWMFRRGKNSHYFDLVFNLEKNTRADATDIENFASRYLYMATLYKYTYNVE 179
Db 233 KRKYVAGQGRINYY-----WTLEPG-----DTIIFANGNLIA 267
Qy 180 FGASFFNKLSTFTGLPGWGIKRALKOIRSNLPIDIGTEHSVRLQHTSS--YKDYMDT 237
Db 268 PWYAFALSRGFGSG-----IITSNMSMDCDPCQIPQGAISLSLFGVHVHV 315
Qy 238 QIPALPKFAKRPRLMV-----QRLATVAGYVDTP--YKKWY----- 274
Db 316 TIGCEPKYVRSYTKLRMTVGLRNIPSIQSRGLGALGFLGGWTGMDIGWGYHNOEG 375
Qy 275 -----MKLNPMVNR-----FIPTKKFFNKEIRPSKALKEKYSTDTKD 314
Db 376 SGYAADQKSTQNAINGITKVNIVIEKMTQTFVAGKERNK-LEKRMENLNKRYVDGFLD 434
Qy 315 -----LEENKIGQGVDFENKEIRDPKALKEKVSNDKADLEENKIGQGVDF-- 362
Db 435 IMTYNAELLVLEEN---ERTLDFHDSNKNLYEKVKSOLKNNKE-----IGNGCFEYH 486
Qy 363 -INNEIRDPKALIRKYSTGAEJL-----FENKIGQGVDFINNE 401
Db 487 KCNNE-----CMESVKNQTYDPRYSESKLNREKIDGVKLE 523

RESULT 20
YB47_METJA
ID YB47_METJA STANDARD: PRT: 462 AA.
AC Q58547;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1147.
GN MJ1147.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
RA Uitterlbeck T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
```


RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii.";
RL Science 273:1056-1073(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67557; AAB9156.1; -.
DR TIGR: Mj1147; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 381 401 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
SQ SEQUENCE 462 AA: 54433 MW: 91608862971776 CRC64;

Query Match 4.6%; Score 107.5; DB 1; Length 462;
Best Local Similarity 22.7%; Pred. No. 3.8;
Matches 44; Conservative 30; Mismatches 75; Indels 45; Gaps 7;

QY 130 KNHSFHLVNLKENTROAD-----ATDIENFASRYLMATLYKTYTNVDEGASF 184
DB 135 KSHVEFDISFLSFTDCRRDDVSFKKIDENKGLAIFLKTQFLKKHTTENFOLS- 253
QY 185 FNKLSFTTGLFGWGIKRALKOIIRSNLPDICTE---HSVSRLOHTTSYKYDMDTOIP 240
DB 254 --KISF-----LKTIDREVLLCDVYKKEILSHKIRIREDSGNKKDKLENLX 299
QY 241 ALPFAKPSLAVVORLLATVAGYVDTPWYKWKMLKNPV---NRFLEPTKFFENK 296
DB 300 ELGLSYRY-----IIDQFNYSVLAEYRNLRISINNNTYEASMLYME 345
QY 297 ---IREPSKALKEK 307
DB 346 MELIKFSNGRPEK 359

RESULT 21
HEMA_IJAH3 STANDARD: PRT: 550 AA.
AC P11134; Q84025; Q84026;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin [contains: Hemagglutinin HA1 chain; Hemagglutinin HA2
DE chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Swine/Hong Kong/136/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11498;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88101364; PubMed=3336940;
RX Kida H., Shortridge K.F., Webster R.G.;
RT "Origin of the hemagglutinin gene of H3N2 Influenza viruses from pigs
RT in China.";
RL Virology 162:160-166(1988).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19056; AAA43211.1; ALT_TERM.
DR PIR: A29971; HMV52.
DR HSSP; P03437; ZHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR ProDom; PD000225; Hemagglutn; 1.
KW Hemagglutinin; Envelope protein; Glycoprotein.
FT NOW TERM 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA: 61580 MW: 991608860292452 CRC64;

Query Match 4.5%; Score 107; DB 1; Length 550;
Best Local Similarity 18.4%; Pred. No. 5.1;
Matches 95; Conservative 63; Mismatches 164; Indels 194; Gaps 23;

QY 5 DSVGVYTKTLAASESVDSANAYMINDMSDYLS-----AVSD 43
DB 77 DVFOQETMDLEFERSNARSCYCPY---DVPDYASLSLAVASSGTFTEGFTWGTQTQ 132
QY 44 NFAERICQOVKRGSCMSVSAVYMSRCAKQDCPLQSLKYBLEAKYQPLTLPPDQGLEA 103
DB 133 NGSSACKRGP-----ASGFESRL--NWLTKSGSTPYV---LNVTPMNDNDFKL 177
QY 104 FI-----LFKESDANPANSTK-----RFWRRFRGKNHSYFHL 138
DB 178 YIMGVHHPSTNOEQNLVQASGRVTVSTRSQQTIPINIGSRPVVQSGSRISITW--- 234
QY 139 VFNLEKNTVDADATDIENFASRYLYMATLYKTYTNVDEFGASFENKLSFTTGLFGWG 198
DB 235 -----TIVKPGDVLVINSNGN--LIAPRGYFKMRT-----GKS- 265
QY 199 IKRALKOIIRSNLPDICTENHSVSRLOHTSS--VKQVMDTOIPALPKFAKPSLMV--- 253
DB 266 -----SIRSDAPIDTCVSECTIPNGSIPNDKPPQVNNKITTYGACPRVYKNSLKLATG 319
QY 254 -----VORLLATVAGYVDTPW---YKKWY-----MKLK-----NF 280
DB 320 MRNVPEKQTRGLFGAIFIEIENGWGMIDGWYGFRRHNSSEGTGAADLKSQOALDQING 379
QY 281 MYNRYEFTTKKFFN--KEIREPSKALK--EKVSTDTKDLFENKIGQTVDFENKKEIRDP 335
DB 380 KLNRIEKTNEKHQIEKEFSEVEGRIDLEKYVEDTK-----IDLMS----- 422
QY 336 SKALKEKXVNDAKDLFEKKIGQTVDFINNEIRDSKALLIRKYSVGAEDLENKIGQTV 395
DB 423 -----YNADVLALEN--OHTIDLTDSSENNKLFETKROLRENAADM-----GNGCF 467
QY 396 DFINNEIRDSKALLIRKYTEADLEFNKIGQTV 431
DB 468 -----KTYHKCDNACISIRNGTYD 487

RESULT 22
RRPL_DUGBV STANDARD: PRT: 4036 AA.
AC Q66431;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (L protein).
GN L.
OS Dugbe virus.

OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Nairovirus.
OX NCBI_TaxID=11595;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE ARD 44313;
RX MEDLINE=96332515; PubMed=8760425;
RA Marriott A.C., Nuttall P.A.;
RT "Large RNA segment of Dugbe nairovirus encodes the putative RNA
polymerase."
RL J. Gen. Virol. 77:1775-1780(1996).
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
[RNA] (N).
CC -I- SIMILARITY: CONTAINS 1 OTU DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: 015018; AAB18834.1; .
DR InterPro: IPR003323; OTU.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF02338; OTU.1.
DR PROSITE: PS50802; OTU.1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 1.
DR Transferrase: Nucleotidyltransferase; RNA-directed RNA polymerase.
KW DOMAIN
FT 29 158 OTU.
SQ SEQUENCE 4036 AA; 459382 MW; E2EDF0B435BE31BD CRC64;

Query Match 4.5%; Score 106.5; DB 1; Length 4036;
Best Local Similarity 20.4%; Pred. No. 64;
Matches 75; Conservative 64; Mismatches 141; Indels 87; Gaps 17;

OY 148 TRDADATDI--ENFASRYL-----YMATLYKTYTWNVDEFGASFFNKLSFTTGILG 196
DB 271 TRESEATSIISLQGLKHLTRDRHIMETARSKLYTRDLIDLDIGGLRSSFP---G 326
OY 197 WGIKRAKQIIRSNPLDIDGT-----EHSVRLQIHITSSYKDYMDTOIIPALP 243
DB 327 LGLERYI-QLHSELVLDLVTVYLAVALSTPLGSGNNKKKQFITNC--LMTKLSGR 382
OY 244 KPAKRSLMNVQRLAT---VAGYVTPMYTKKYMKLKFM-----VNRVFTTKKFFNK 295
DB 383 VF-KALSKLTGQALYTPKRAVSIVSOELYGLKMLVKNNLECMGPISMLALNLFNDM 441
OY 296 EIREPSKALKEKYSTDTKDLFENKIGGTVDENKEIRD-----PSKALKEKVSNDK 348
DB 442 QLODYILEMSEKIDNSD-----VEYTHREISDLHLTLVERLSKLQKSDYNEIK 491
OY 349 DLF-----ENKIGQTVDFINNEIRDPKSKALIRKVSSTAEDLFENKIGQTVDFIN 399
DB 492 LMEKEEVLTKRSQSYGNFEEFLINDYFK--KKDIMKPVSTGSKASTGIG--NVLSYAH 548
OY 400 N-----EIRDSKAL-----IRKVTYEAADLFENKIGQTVDFINKEIRDPKSKALIRK 447
DB 549 NYLSKESLRMTSEDTOLLIEIRKLKLOGDLSEPPVA-----IICDKLEDFRKLFRE 603
OY 448 VSTEADN 454
DB 604 LPEECSS 610

RESULT 23
HEMA_IADH6
ID HEMA_IADH6 STANDARD: PRF: 550 AA.
AC P12587; 084017;
DT 01-OCT-1988 (rel. 12, Created)
DT 01-APR-1990 (rel. 14, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain];

DE Hemagglutinin HA2 chain] (Fragment).
GN HA
OS Influenza A virus (strain A/Duck/Hokkaido/9/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11362;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
ducks".
RL Virology 159:109-119(1987).
CC -I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -I- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL: M16742; AAA43148.1; .
DR PIR: F27813; HMIY98.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin.1.
DR ProDom: PD000225; Hemagglutn.1.
KW Envelope protein; Hemagglutinin; glycoprotein.
FT NON_TER 1 1
FT CHAIN 1 328
FT CARBOHYD 330 550
FT CARBOHYD 22 22
FT CARBOHYD 38 38
FT CARBOHYD 165 165
FT CARBOHYD 285 285
FT CARBOHYD 483 483
FT CONFLICT 8 8
SQ SEQUENCE 550 AA; 61711 MW; 67BCD85F44736CFE CRC64;

Query Match 4.5%; Score 106; DB 1; Length 550;
Best Local Similarity 18.3%; Pred. No. 5.9;
Matches 92; Conservative 65; Mismatches 153; Indels 192; Gaps 23;

OY 19 ESDVSAANAMVMSDMSDIYS-----AVSDFARISQVPPKGS 57
DB 89 ERSNAFSNCPY--DVPDVAISLSLVAASGTLFEITGFTWGTQVTONGSGNACKRGPB- 144
OY 58 NCSASVAVMRCADKDCLOSLKYPLEAKYOPLTLPDYOLEAFI----- 105
DB 145 -----SGFSSRL--NWLTKSGSTYV---LNVTPNPNNDNDKLYIMVHHPTNOEQ 191
OY 106 -LFKESDANPANSTK-----RFMFRFRGKNHSYFHDVLVFNLEKNVTRDAD 152
DB 192 TNLVQASGCVTVSTRSOOTIIPNIGSRPWRVQSGRISYIW-----TVVAPGD 241
OY 153 ATDIENPASRYLYMATLYYTYTNNVDEFGASFPNKLSFTTGLGCGWGIKRAKQIINSNLP 212
DB 242 VLIVINSNGN--LIAPKGYEMRT-----GKS-----SIMSDAP 273
OY 213 LDIGTEHSVRLQHTSS--YKDYMDTOIIPALPKFAKFSILAV-----VORLLA 259
DB 274 IDICISECITPNSGIPRDKRFQVNNKITTYGACRYKQNTLKLATGMRNVPKQGTGLG 333
OY 260 TVAGYVDPW---YKKWY-----MKLK-----NEMVNVRFIPTKKFFN 294
DB 334 AIAGFIENGEWGMIDGWYGRHONSECTGOAALDKSTQAAIDQINGKLNIVIEKTEKFFH 393

```

OY 235 ---KEIREPSKALK--EKVSTDTKDLFENKIGQTVDFPNKEIRDPKALKKEKVSNDAKD 349
DB 334 QLEKEFSEVEGRIDLEKEKVEDTK-----IDLMS-----YNAADVLY 429
OY 350 LEENKIGQTVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQTVDFINNEIRDPKALK 409
DB 430 ALEN---QHTIDLTSEMKKLFERTRRQLRENAEDM-----GNGCF----- 467
OY 410 IRKYVTEADDLFENKIGQTVDF 431
DB 468 --KIYKCDNVCIESIRNGTYD 487

RESULT 24
HEMA_IADH7 STANDARD: PRT: 550 AA.
ID HEMA_IADH7 084018: 089470:
AC P12588: 084018: 089470:
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor (Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAI chain) (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/10/85).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OX NCBI_TaxID=11363;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawoka Y., Naeye C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks."
RL Virology 159:109-119(1987).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M16743: AAA43149.1: -.
DR PIR: G27813: HMIY15.
DR InterPro: IPR001364: Hemagglutn.
DR Pfam: PF00509: Hemagglutinin; 1.
DR Prodom: PD000225: Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
FT CHAIN 1 330 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 550 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 550 AA: 61761 MW: 66E81793281D53EB CXC64:

Query Match 4.5%; Score 106; DB 1; Length 550;
Best Local Similarity 18.3%; Pred. No. 5.9;
Matches 92; Conservative 65; Mismatches 153; Indels 192; Gaps 23;
OY 19 ESDVSAANAYMINSDMSYLS-----AVSDNFAERICSGVPGKS 57
DB 89 ERSNAFSCYPI--DVPIYASLRLSVASSGTLFTTEGTVGTQVNGSGNACKRGPN-- 144

```

```

OY 58 NCSASVAYSMSRCANODCLTIQSLKYLEAKYQPLTLPDPOLEAFT----- 105
DB 145 -----SCFESRL--NMLTSGSTYPV-----LNTWPMNDNFDKLYIGVHHPSNQO 191
OY 106 ---LFKESDANPANSTK-----RPMRFRGRKHSFPHDLYVFLKKNYTRQAD 152
DB 192 TMLLYVQASGRVTVSTRRSOQTIIPNIGSRWVGSGRSTIYW-----TVVKPD 241
OY 153 ATDIENFASRYLYMATLYKYTYTNDEFGASPFNKLSFTTGLGWMGIRKALQIIRSNLP 212
DB 242 VLVINSGN--LIAPRGYKMT-----GKS-----SIRMSDAP 273
OY 213 LDIGTEHVSRLQHTSS--YKDYMDQIPLPKPAKPSLMV-----VORLLA 259
DB 274 IDTCISECTTPNGSIPNDKPFQNVNKITYGACPKYKQNTLKLATGMNVPEKOTRGLFG 333
OY 260 TVAGYVDTPW---YKKWY-----MKL-----NPMVARYPLPKKFFN 294
DB 334 ATAGFLENMCEGMIDMGWGFRRQNSEGTGAADLSTQAIDQINKLNRVLEKTEKPH 393
OY 295 ---KEIREPSKALK--EKVSTDTKDLFENKIGQTVDFPNKEIRDPKALKKEKVSNDAKD 349
DB 394 QLEKEFSEVEGRIDLEKEKVEDTK-----IDLMS-----YNAADVLY 429
OY 350 LEENKIGQTVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQTVDFINNEIRDPKALK 409
DB 430 ALEN---QHTIDLTSEMKKLFERTRRQLRENAEDM-----GNGCF----- 467
OY 410 IRKYVTEADDLFENKIGQTVDF 431
DB 468 --KIYKCDNVCIESIRNGTYD 487

RESULT 25
YMO8_YEAST STANDARD: PRT: 624 AA.
ID YMO8_YEAST
AC P54074:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 71.4 kDa protein in ILV2-ABE17 intergenic region.
GN YMR119W OR YMR718.18 OR YMR564.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-431 FROM N.A.
RA STRAIN-S288C / AB972;
RC Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 345-624 FROM N.A.
RA STRAIN-S288C / AB972;
RC Lye G., Churher C.M., Barrell B.G., Rajandream M.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO YEAST YNL008C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: 249702: CAA89757.1: -.
DR EMBL: 249273: CAA89268.1: -.
DR SCD: S0004725: YMR119W.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 32 52 POTENTIAL.

```

FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 274 294 POTENTIAL.
 SQ SEQUENCE 624 AA: 71443 MW: E2F5149F89A09655 CRC64;

Query Match 4.5%; Score 106; DB 1; Length 624;
 Best Local Similarity 25.6%; Pred. No. 6.9;
 Matches 68; Conservative 43; Mismatches 87; Indels 60; Gaps 17;

QY 171 YKTYNVDVEFGA---SFFNKLSFTTGLFGWGJKRALKQIIRSNLPDICTEHSV-SRLQH 226
 DB 321 YLNTYGEEDFSAMATKFFALLCSGTLMKGIIRREPAI---NIPNNVEKFFISGLY 377
 QY 227 TTSSTKDYDIOIPLPK-----FAKPSLAVVQRLATVAGYVDTPTWKYKKWIKLNF 281
 DB 378 LSKPYKE--NTSI-SFFKNSSILKORFIMPEKSIIT-----WMKK---L 417
 QY 282 VNRVFIPTKFFPKET--REPSKALKEVSTDTKDLFENKIGQ-----GTVDFFNK 330
 DB 418 VQVFFGFNDKDEIDPDNDPSKMLK---ITKTNLS--NNSAGHKEDIELELNTSDEYS 473
 QY 331 EIRDSPSKALKEVSNDAKDLFENKIGQGTVDFINNEIRDPKALIRKYSTGAEDLFENKI 390
 DB 474 EDERSE-VESLGDSDEENLEEDSL-----IFNETRD--ALL-----DLFSSD 514
 QY 391 GGGTVDFINNEIRDPKALIRKYITE 416
 DB 515 NEVHTDY--NWIMSTSRILQOKRLSD 538

RESULT 26
 HEMA_IADHK STANDARD: PRT: 550 AA.
 AC P43257;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 Hemagglutinin HA2 chain] (Fragment).
 GN HA.
 OS Influenza A virus (strain A/Duck/Hong Kong/7/75).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 OX NCBI_TaxID=11364;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91341491; PubMed=1875195;
 RA Yasuda J., Shortridge K.F., Shimizu Y., Kida H.;
 RT "Molecular evidence for a role of domestic ducks in the introduction
 of avian H3 influenza viruses to pigs in southern China, where the
 A/Hong Kong/68 (H3N2) strain emerged.";
 RT J. Gen. Virol. 72:2007-2010(1991).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC or send an email to license@sib-sib.ch).
 CC EMBL: D00929; BAA00769.1; -
 DR HSSP: P03437; SHMG.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR ProDom: PD000225; Hemagglutin; 1.

KW Envelope protein; Hemagglutinin; glycoprotein.
 FT NON_TER 1 1
 FT CHAIN 1 328
 FT CHAIN 330 550
 FT CARBOHYD 8 8
 FT CARBOHYD 22 22
 FT CARBOHYD 38 38
 FT CARBOHYD 165 165
 FT CARBOHYD 285 285
 FT CARBOHYD 483 483
 SQ SEQUENCE 550 AA: 61549 MW: 864639829FEI8A9 CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 550;
 Best Local Similarity 18.0%; Pred. No. 6.4;
 Matches 88; Conservative 69; Mismatches 154; Indels 177; Gaps 22;

QY 12 KTLIASEVDSANANAYMINSMDYLSAVSDNFARICSOVQKSGNSCASVAVSRCA 71
 DB 109 RSLVASSGTLFTTBFT-----WTGVTONGSSACKRGF-----ASGFSRL- 151
 QY 72 KODCLTLQSLKYPLEAKYPLTLPDYQLDAFT-----LFKESDANPANS 117
 DB 152 --NMLTKSGSTYFV---LNVTPNNDNFDKLYIMGVHHPSTNQEQTNLYVQASGRVTS 205
 QY 118 TEK-----REWMRFRRGKNHSYFHDLYFNLEKNVTRDADATDIENASRYLYM 166
 DB 206 TRRSQGTIIPNIGSRPVRVQSGRISYV-----TIYPGVLYVNSGN--LIA 253
 QY 167 ATLYKYTYNVDVEFGASFENKLSFTTGLFGWGJKRALKQIIRSNLPDICTEHSVRLQH 226
 DB 254 PRGYFKMRT-----GKS-----SIMRSDAIPIDCVSCTIPNPS 287
 QY 227 TTS--YKDYMDIOIPLPKFAKPSLAVV-----VQRLATVAGYDTPW---Y 270
 DB 288 IPNDKPFQVNVKITYGACPKRYVQNSLKLATGARNVPEKOTRLFAIGAFIENGEGMI 347
 QY 271 KKYV-----MLK-----NEMVNVFIPTKFFN---KEIRPSKALK 305
 DB 348 DGYGFRHNSSEGTGAADLKSTQAAIDQINGLNRYIKTKNEKFOIEKSEVSGRIQ 407
 QY 306 --EKVSTDKDLFENKIGQGTVDFFNKIEIRDPKALKEVSNDAKDLFENKIGQGTVD 363
 DB 408 DLEKYVEDRK-----IDLMS-----YNADVLALEN--QHTIDLT 440
 QY 364 NNEIRDPKALIRKYSTGAEDLFENKIGQGTVDFINNEIRDPKALIRKYITEADLFEN 423
 DB 441 DSEMNKLFKTRQRLRENAEDM-----GNCCF-----KIYHCKDNACIE 479
 QY 424 KIGQGTVD 431
 DB 480 STRNGTYD 487

RESULT 27
 HEMA_IAEN7 STANDARD: PRT: 566 AA.
 AC P03440;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
 Hemagglutinin HA2 chain].
 GN HA.
 OS Influenza A virus (strain A/England/321/77).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 OC Influenza virus A and B group; Influenza A viruses;
 OC Influenza A virus.
 OX NCBI_TaxID=11378;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83110955; PubMed=6822816;
 RA Hauptmann R., Clarke L.D., Mountford R.C., Bachmayer H., Almond J.W.;
 RT "Nucleotide sequence of the haemagglutinin gene of influenza virus

RT A/England/321/77.":
RL J. Gen. Virol. 64:215-220(1983).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05907; CAA29337.1; -.
DR PIR: A04052; HMYV6.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR PRODOM: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 346 566 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 566 AA; 63608 MW; FA5B886FF4B8C888 CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 566;
Best Local Similarity 19.1%; Pred. No. 6.6;
Matches 103; Conservative 73; Mismatches 185; Indels 177; Gaps 27;

OY 11 TTTLLAASESVDSANNAVMSINDMSDYLSA---VSDNFAERICS---OYPKSNCASASVS 64
DB 26 TATLCLAHNAVNGTLVTTTNDQIEVNTATLVOSSSTGRICDSPHRLDKNCCTL-ID 84
OY 65 AVMSCARQDCITLQSLKLYPLEAKYO-----PLTLPDPYOL-----EAAFILEKESDA 112
DB 85 ALL---GDPHCGFQNEKMDLFEVRSKAFSNCCYPDVPDYASRLVASSGLEFTEGFC 141
OY 113 NPANSTKRFMMRRFRGKNHSYFHDLVFNLEK-----NVRDADATDIENFASRL 164
DB 142 NMTGVTONGSAVACRGRPDNSFFSL--NMLYKSESTYPLVLTMPNN---DNFDKLYI 195
OY 165 Y-----MATLYYKT-----YTNDVEFGA--SFFKKLSFTTGLFGM---- 197
DB 196 MGVHHPSTDKETKLYVASGGRVYSTRKSQOTIIPNVSRRWVAGLSRISIT-YTIYK 254
OY 198 -----GIRALKQIIRSNLPDIDIGTEHSVRLQHTSS--YKDYM 235
DB 255 PGDILLINSNGNLIPRGYFKIRTKSSIMRSDAPIGTCSSECIIPNCSIPNDKDFQVNV 314
OY 236 DTQIPALPKAFKRFSLM-----VQRLATVAQYVDTPM---YKKKY----- 274
DB 315 KITYGACPRYKQNTLKLATGKRNVPKQTRIGFALIGFLENGEMGIDWYGRHONS 374
OY 275 -----MKLK-----NFAVNRVFIPTKKFEN---KEIREPKALK--EKVSTDRK- 313
DB 375 EGTGAADLKSTQALIDINGKLNIVIEKTNKEFHQIEKESFEVSGRIODELEKYVEDRKI 434
OY 314 DLFEKKIGQGVDFPNKEIRDSKALKKEVSNDAKDLFEKKIGQGVDFINNEIRDSPSKA 373
DB 435 DLMS-----YNAEL---LVALE-----NQHTIDTIDSEMKKLEFEK 466
OY 374 LIRKSTGAEDLFEKKIGQGVDFINNEIRDSKALKIRKYTEADLFEKKIGQGVDF 431

DB 467 TRROLENAEDM-----GNGCF-----KIYKCDNACIGSRNGTYD 503
RESULT 28
ID Y075_MYCPN STANDARD; PRT: 1030 AA.
AC P75556;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG075 homolog (G07_Orfi1030).
GN MPN213 OR MP618.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE-97105885; PubMed-8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000059; AAB96266.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
SQ SEQUENCE 1030 AA; 116017 MW; DAA4E48035973E07C CRC64;

Query Match 4.5%; Score 105.5; DB 1; Length 1030;
Best Local Similarity 21.0%; Pred. No. 14;
Matches 92; Conservative 71; Mismatches 171; Indels 105; Gaps 21;

OY 78 LOSLKYPLEAKYOP-----LTLPD-----PYOLEAFILEKE-----SD 111
DB 51 LMSVKLP---KAQPAATRTIVENGTDKLVNKKSSPOQLFLAKNALDKLOCEPDKFLSD 107
OY 112 AN--PANSTKRFMMRRFRGKNHSYFHDLVFNLEKNTVRDADATDIENFASRYLYMATL 169
DB 108 AKAFAPLATADLOEWVDOQLFNPDSFFDL--SAPRSNFTLSSDKRASIDFIFR----- 158
OY 170 YKTYTNVDEFGASFNNKLSFTTGLFGMGIRALKQIIRSNLPDIDIGTEHSVRLQHTS 229
DB 159 ----FTN-----FTESVQLKLKPGVGVVSDSKSPFYVYVNASQKLLVPL 201
OY 230 SYKD-----YMDQIPALPKFAKRFSLMVYQRLATVAQYVDTPMYKKWKLKLFMVN 283
DB 202 SLDPDTGLGNTMYFHDITLNGKVNKFSFNPFTNLNLA-----FSNVYGVGVDFEQA 253
OY 284 RYFIPTKKFENKEIREPSKALKKEYSTDKDLFE-----NKIGQGVDFPNKEIRDSK 337
DB 254 KNLVKGKRYLNTNVR--MEDVKKDVNANKNQFDIAKIIAELMGKALKEFGNQDGPPLS 311
OY 338 ALK--EKVSNDAKDLFEKKIGQGVDFINNEIRDSKALKIRKYSTGAEDLFEKKIGQGV 395
DB 312 FLKVDKQKDEDEKLF-NLVRPGLKFPYKDLIOSSQA--EKKITVYKLIIFPNK---KTI 365
OY 366 DEINNEIRDP-----SKALIRKY---TEADLFEKKIGQGVDFINKEIRDSKA----- 443
DB 366 LNLKELSIPELNSSLGLVDFLFDGTFDSQGLYERL--QSFKDLIVPAVKTNKEKTAALSP 423
OY 444 LIRKYSTEAD-----NLLEK 458


```

Db 423 ---IYKLRRIESTEDKDYETLNDKIESLEKOLDKONLNRSLRISKIDETTORYL 479
Oy 125 RRRGKNHSYFDLVNLEKLVNRDADATDIENFASRYLYATLYKTYTNDVDEGASF 184
Db 480 DAK-----DIELKQSVSIRKQETLAKMNAQLELSAIIQOT-----515
Oy 185 FFKLSFTTGLFGWGIKRALKQIIRSNLPIDIGTSHS---VSRLQHTSSYKDYMDTOJP 240
Db 516 VVELTAT-----KNELSQVATTNQSLFANEESKVLLEGTLAIFDSFYQIIMQIEKP 567
Oy 241 ALPRFAKRPSLWVQORLLATVAGYDTPWYKKWYMKLKFMMNR-----284
Db 568 DYVPIJSK-----POLTNQESIT-OTDIKQMLQKLSLNADYANLQSVSELSDKS 619
Oy 285 --VFIPTKKFKFNKEIRPSKALKEKSTDTKDLFENKIGQTVDFENKEIRPSKALKEK 342
Db 620 QIISIVPRIVIRILKENKAVENNA---ELLESVAATSAVSAL---VOQYERSEKON 673
Oy 343 VS-----NDAKDLFENKIGQTVDF--INNEI-RDPSKALIRKVSAGADLFENKI 390
Db 674 VKLRQEFELKLDLQRLLE---ONQOTDESISEFISRPD--AFNNRLN---DERFONLR 724
Oy 391 GCGTVDFTINERDPSKALIRKYVTEADDLFENKIGQTVDFINKEIRDPSKALIRKVS 450
Db 725 QO--YDMSKTSALETTKIKEMESTADQAVKSEMSK-----LNTQL-DELSLFFVKYR 776
Oy 451 EADNILE 457
Db 777 KADIFE 783

RESULT 31
AC NFM_CHICK STANDARD: PRT; 857 AA.
ID P16053;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NFM.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90174973; PubMed=2106668;
RA Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
RT "Isolation of the chicken middle-molecular weight neurofilament
RT (NF-M) gene and characterization of its promoter.";
RL Nucleic Acids Res. 18:521-529(1990).
RN [2]
RP SEQUENCE OF 259-857 FROM N.A.
RX MEDLINE=88112814; PubMed=3123320;
RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D., Betz H.;
RT "Identification of gene products expressed in the developing chick
RT visual system: characterization of a middle-molecular-weight
RT neurofilament cDNA.";
RL Genes Dev. 1:699-708(1987).
RN [3]
RP FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
RP AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
RP -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
RP PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
RP THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
RP INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
RP OF AXONAL CALIBER.
RP -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
RP OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
RP LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
RP CONCURRENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
RP -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17102; CAA34958.1; -
DR EMBL: X05558; CAA29073.1; -
DR PIR: A27040; A27040.
DR PIR: S08061; S08061.
DR PIR: S15762; S15762.
DR InterPro: IPR001664; IF.
DR Pfam: PF00038; filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
KW Glycoprotein.
FT INIT_MET 0
FT DOMAIN 1 98 HEAD.
FT DOMAIN 99 406 ROD.
FT DOMAIN 407 857 TAIL.
FT DOMAIN 99 130 COIL_1A.
FT DOMAIN 131 143 COIL_1B.
FT DOMAIN 144 242 COIL_1B.
FT DOMAIN 243 259 LINKER_12.
FT DOMAIN 260 281 COIL_2A.
FT DOMAIN 282 285 LINKER_2.
FT DOMAIN 286 406 COIL_2B.
FT CARBOHYD 46 46 O-LINKED (GICNAC) (BY SIMILARITY).
FT CARBOHYD 426 426 O-LINKED (GICNAC) (BY SIMILARITY).
FT CONFLICT 546 546 G -> R (IN REF. 2).
SQ SEQUENCE 857 AA; 95704 MW; 4E2E0FC6AC64778B CRC64;

Query Match 4.5%; Score 105; DB 1; Length 857;
Best Local Similarity 18.1%; Pred. No. 12;
Matches 94; Conservative 78; Mismatches 183; Indels 164; Gaps 20;

Oy 2 APSDSVGDVTKTLAASEVSD-----SAANAAYINSDMSDYLSAVSDNFA---ERIC 50
Db 58 APRTAG---STVLSAEELDVQSOSLLNGAAELKRSRNEKEQGLGDRFAGYIEKV- 113
Oy 51 SQVPGKSCSASVAYMSCAKODCLTLOSILYPLEAKYORPLTPDYOLEAFILFKES 110
Db 114 -----HYLEDQNKIEIEALALROKHNAGRAQ---LGDVDEQ-----147
Oy 111 DANPANSTKRRFMRRRRKNHSYFDLVNLEKLVNRDAD--ATDIENFASRYLYMAT 168
Db 148 -----LRELKALDEOVSH-----KAOIQDSEHIEDIQRLEREFDEAR 188
Oy 169 LYKTYTNY-----DEFGASFNKLSTFTGLFGWGIKRALKQIIRSNL 211
Db 189 LNDPEATIALRKEMERASLMRAELDKKVQSLQDEVAFLRGNHEEVALLAQ-----243
Oy 212 PLDIGTSHSVRLQHTSSYKDYMDTOI-PALPRFAKRPSLWVQORLLATVAGYDTPWY 270
Db 244 -----QASHATYERKDYLTDLTTALKEIRAQLECSQDHNM-----HQAEWF 286
Oy 271 KKWYMKLKNFMVNRVFIPPKFFENKEIRPSKALKEKSTDTKDLFENKIG--OGTVD 326
Db 287 KCRYAKLTBEA-----EONKEATRSAAKEIEIYERQLOSKSIELESVRGTRKE 333
Oy 327 FENKEIRDPSKALKEKVSNDADLFENKIGQTVDFINNEIR-----368
Db 334 SLERQISD-----IEBHNNDLTTY-----QDTIHLELNGITKEMARHLREYODLLN 383
Oy 369 -----DPSKALIRKVSAGADLFENKIGQTVDFINNEIRDPSKAL---IRKYVTEAD 419
Db 384 VKMALDIEIAARKLKEGETRFSAFSGSITGPIFTH--RQPSVTIAYSTKIQTIEPK 441
Oy 420 L-----FENKIGQTVDFINKEINDPSKALIRKYSTEA 452

```


DB 442 LKYNHKEVEIEETKEVEDEKSEMEDALSAIAEEMAKA 480

RESULT 32

YF05_METUA STANDARD; PRT; 778 AA.

AC 058900;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative ATP-dependent RNA helicase MJ1505.

GN MJ1505.

OS Methanococcus jannaschii.

OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;

OC Methanococcus.

OX NCBI_TaxID=2190;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;

RX MEDLINE=96337999; PubMed=8688087;

RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,

RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,

RA Kierlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,

RA Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glodek A.,

RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,

RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,

RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,

RA Klenk H.-P., Fraser C.M., Smith H.O., Woese G.R., Venter J.C.;

RT Complete genome sequence of the methanogenic archaeon, Methanococcus

RT jannaschii."

RL Science 273:1058-1073(1996).

CC -! SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD

CC SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: U67591; AAB99518.1; -

DR TIGR: MJ1505; -

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR002464; DEAD_ATP_helicase.

DR InterPro: IPR004016; ERCC4.

DR InterPro: IPR000445; HHH.

DR InterPro: IPR003583; HHH.1.

DR InterPro: IPR001650; Helicase_C.

DR Pfam: PF00270; DEAD; 1.

DR Pfam: PF02732; ERCC4; 1.

DR Pfam: PF00271; helicase_C; 1.

DR Pfam: PF00633; HHH; 1.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELICC; 1.

DR SMART: SM00278; HHH1; 1.

DR PROSITE: PS00690; DEAD_ATP_HELICASE; FALSE_NEG.

DR Hypothetical protein; ATP-binding; RNA-binding; Helicase;

KW Complete proteome.

KW NP_BIND 35 42 ATP (POTENTIAL).

FT SITE 137 140 DEAD BOX.

FT DOMAIN 532 535 POLY-GLU.

SQ SEQUENCE 778 AA; 88882 MW; DCBIF2208013368 CRC64;

Query Match 4.4%; Score 104.5; DB 1; Length 778;

Best Local Similarity 18.5%; Pred. No. 11;

Matches 103; Conservative 91; Mismatches 163; Indels 199; Gaps 28;

OY 5 DSVGDVTKTLIASESVDASNAANVIMNSDKSDYLSAVSNFAEIRCSQYKSGNSCASVS 64

DB 242 NSIADVTKELE-----LNKKLSYDEEVKVELI-KVCEALKLMHARELLE 288

OY 65 A-----YMSRCADCLTLOSLEAKYQPLTPDPYOLEAFLFKESDANPAN 116

DB 289 SGGKSVFLANIKLSMQRTSAKSI-----VDEKVEAVNLMLGDVEHP-- 334

OY 117 STEKRFMRFRCKNSYFDLVNLEKAVTDADATDIENFASRLVATLYKTYTN 176

DB 335 -----KLGK-----VVDWVKNLKKN-KD-----ERIIIPAQ-YRDTVER 367

OY 177 VDEFGASFNNKLFPTTGLPFGWGIKRALQIIRSLPLDICTEHSVSLQHTSSYKD-- 233

DB 368 I-----VNLTLQNGIKAIRITGQANKEGK 392

OY 234 -YMDQIPALPKFKRESLWY-----VORLLAVAG 263

DB 393 MSQKEQLEALERFKKESVAVTSVSEGDIPSVNIIIEPVPSEIRIQGRMRMG 452

OY 264 -----YDTPWYKKWYK-LKN-FWYNRVPIPKKFPK---EIRE 299

DB 453 EGGKVVYLIAGTADENAYRSALYKEREMRLKMKCYLLNKRLL--OKKFEKSEBIKE 510

OY 300 PSKALKK-----VSTDKDLFENKIGGV-DFPNK-EIRDPKALKKESNDK-- 348

DB 511 ETEELKEKTESYAVKEETKEDEKTKPVTIDFKOLEVERKSSSEDKIKQETKIP 570

OY 349 -----DLFENKIGGVDFINNEIRD-----PSKALRKYSTGAEDLFENKIGQ 392

DB 571 KKPKIIVDRKKNMKLHNANIELKTEVGVDYVSDVYVERKT--ADEPVNSIIDK 628

OY 393 GTVDFINNEIRDSKALIRRVYTEADDLFENKIGGV----DF-----INKEIRDPK 442

DB 629 RLFSQALKN-LKVEKPLIVEGENFSRLHENAL-KGALISIIIDFGIPIIFTKNAEETAD 686

OY 443 ALIRKYSTENDNLEK 458

DB 687 LLIR--IAEKQIRK 700

RESULT 33

HEMA_IADH1 STANDARD; PRT; 550 AA.

ID HEMA_IADH1

AC P12582; Q84021; Q84022;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;

DE Hemagglutinin HA2 chain] (Fragment).

GN HA.

OS Influenza A virus (strain A/Duck/Hokkaido/5/77).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza virus A and B group; Influenza A viruses;

OC Influenza A virus;

OX NCBI_TaxID=11357;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87265458; PubMed=2440178;

RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;

RT "Antigenic and genetic conservation of H3 influenza virus in wild

RT ducks."

RL Virology 159:109-119(1987).

CC -! FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO

CC CELL RECEPTORS AND FOR INITIATING INFECTION.

CC -! SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS

CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.

CC -! SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

EMBL: M16737; AAA43143.1; -
 DR PIR: A27813; HMI177.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR Prodom: PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 550 AA; 61705 MW; 7E7ACEF716FC969A CRC64;

Query Match 4.4%; Score 104; DB 1; Length 550;
 Best Local Similarity 18.1%; Pred. No. 8;
 Matches 91; Conservative 65; Mismatches 154; Indels 192; Gaps 23;

19 ESDSAAANAYMINSMDYLS-----AVSDNFAERICSOVPGKS 57
 DB ESNMFSNCPY--DVPDVALSLVASSGTLFTTEGTTWGTGONGSKACKRGP--- 143
 OY 58 NCSASVSAVMSRCANQDCITLQSLKYPLEAKYQPLTLPDYLEAFT----- 105
 DB 144 -----ASGFFSRLL--NMLTKSGSTYPV---LNTVMPNNDNFDKLYLGVGHHPTDQEQ 191
 OY 106 --LFKESDANPNANSTREK-----RPMRFRGKNHSEFHDLYVFLLEKNYTRDAD 152
 DB 192 TLLVYQASGRVTVSTRSQOTIIPNIGSRPWRVGGSGRISTW-----TIVKPGD 241
 OY 153 AADIENFASRYLYMAVLYKYTYTNVDEFGASFENKLSFTTGLFGWIKRALKQIIRSNLP 212
 DB 242 VLVINSNGN--LIAPRGVYKMT-----GKS-----SIMSDAP 273
 OY 213 LDIGIEHSYRLQHTTSS--YKDYMDQIPLPKFAKRSIMV-----VORLLA 259
 DB 274 IDTCISECTTPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGMRNVPEKQTRGLFG 333
 OY 260 TVAGYVDFPW---YKKWY-----MKLK-----NPMVNRVEIPTKKFFN 294
 DB 334 AIAGTIENGMEGIDGWGFRHONSEGTQAADLKSTQAIQINGKLKRVLEKTEKHEH 393
 OY 295 ---KEIREPSKALK--EKVSTDTKDLFENKIGQGVDFENKEIRDPKALKKEKVSNDAD 349
 DB 394 QKEEFSEVEGRIDLEKVEDTK-----IDLMS-----YNDVLY 429
 OY 350 LFEKNIIGQGVDFINNEIRDPKALKIRKYSTGAEDLFENKIGQGVDFINNEIRDPKAL 409
 DB 430 ALEN---QHTIDLTSEENKLFEEKRQRLRENAEDM-----GNGCF----- 467
 OY 410 IRKYTTEADDFENKIGQGVDF 431
 DB 468 --KIYHKCDNACVESIRNGTYD 487

RESULT 34
 HEMA_IADH2
 ID HEMA_IADH2 STANDARD: PRT: 550 AA.
 AC P12583; 084011;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
 DE Hemagglutinin HAI chain (Fragment).]
 GN HA.
 OS Influenza A virus (strain A/Duck/Hokkaido/8/80).
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
 CC Influenza virus A and B group; Influenza A viruses;
 CC Influenza A virus.
 OK NCBI_TaxID=11358;

SEQUENCE FROM N.A.
 RP MEDLINE=87265458; PubMed=2440178;
 RA Kida H., Kawoka Y., Naeve C.W., Webster R.G.;
 RT "Antigenic and genetic conservation of H3 influenza virus in wild
 RT ducks";
 RL Virology 159:109-119(1987).
 CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
 CC CELL RECEPTORS AND FOR INITIATING INFECTION.
 CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
 CC (HAI AND HAZ) LINKED BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: M16738; AAA43144.1; -
 DR PIR: B27813; HMI180.
 DR InterPro: IPR001364; Hemagglutn.
 DR Pfam: PF00509; Hemagglutinin; 1.
 DR Prodom: PD000225; Hemagglutn; 1.
 KM Envelope protein; Hemagglutinin; Glycoprotein.
 FT NON_TER 1
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CHAIN 1 328 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 8 550 HEMAGGLUTININ HAI CHAIN.
 FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 137 137 K -> N (IN PIR DATA BANK).
 SQ SEQUENCE 550 AA; 61659 MW; A107033ACC9CC353 CRC64;

Query Match 4.4%; Score 104; DB 1; Length 550;
 Best Local Similarity 18.9%; Pred. No. 8;
 Matches 92; Conservative 63; Mismatches 169; Indels 164; Gaps 22;

19 ESDSAAANAYMINSMDYLS-----AVSDNFAERICSOVPGKS 57
 DB ESNMFSNCPY--DVPDVALSLVASSGTLFTTEGTTWGTGONGSKACKRGP--- 143
 OY 58 NCSASVSAVMSRCANQDCITLQSLKYPLEAKYQPLTLPDYLEAFTLFKESDANPNANS 117
 DB 144 -----ASGFFSRLL--NMLTKSGSTYPV---LNTVMPNNDNFDKLYI----- 179
 OY 118 TEKRFWMRFRGKNHSEFHDLYVFLLEKN-----VTRDADATDIENFASRYLYNA--- 167
 DB 180 -----W-----GVHHPTSQBOETNLVYQASGRVTVSTRSQOTIIPNIGSRPWRVGGSGR 229
 OY 168 -TLVYKYTYTNVDEFGASFENKLSFTTGLFGWIKRALKQIIRSNLPDIDIGIEHSYRLQH 226
 DB 230 ISITYTWIYVPGDVLVINSNGNLIAPRGYF--KMTGKSSIMSDAPIDCVSECTTPNGS 287
 OY 227 ITSS--YKDYMDQIPLPKFAKRSIMV-----VORLATVAGYVDFPW---Y 270
 DB 288 IPNDKPFQNVNKITYGACPKYVKQNTLKLATGMRNVPEKQAGALGAGIENGMEGNI 347
 OY 271 KKWY-----MKLK-----NPMVNRVEIPTKKFFN---KEIREPSKALK 305
 DB 348 DQWYGFRRHONSEGTQAADLKSTQAIQINGKLKRVLEKTEKHEKESVEYGRQ 407
 OY 306 --EKVSTDTKDLFENKIGQGVDFENKEIRDPKALKKEKVSNDADLFENKIGQGVDFI 363
 DB 408 DLEKVEDTK-----IDLMS-----YNDVLYALEN---QHTIDLT 440
 OY 364 NNEIRDPKALKIRKYSTGAEDLFENKIGQGVDFINNEIRDPKALKIRKYTTEADDFEN 423

```

DB 441 DSEMNKLFETRRLRENAEDM-----GNGCF-----KIYHKCNACIE 479
QY 424 KIGCGTVD 431
DB 480 SIRNGTYD 487

RESULT 35
HEMA_IADH3
ID HEMA_IADH3 STANDARD; PRT; 550 AA.
PI2584; Q84012; Q89793;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HA1 chain;
DE Hemagglutinin HA2 chain] (Fragment).
GN HA.
OS Influenza A virus (strain A/Duck/Hokkaido/33/80).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxId=11359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87265458; PubMed=2440178;
RA Kida H., Kawaoaka Y., Naeve C.W., Webster R.G.;
RT "Antigenic and genetic conservation of H3 influenza virus in wild
RT ducks.";
RL Virology 159:109-119(1987).
CC -I- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -I- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -I- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M16739; AAA43145.1; -.
DR PIR: C27813; HMTV33.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein.
FT NON_TER 1
FT CHAIN 1 328 HEMAGGLUTININ HA1 CHAIN.
FT CHAIN 330 550 HEMAGGLUTININ HA2 CHAIN.
FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 165 165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 483 483 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 550 AA; 61577 MW; 6C30BF67CFDC87DE CRC64;

Query Match 4.4%; Score 104; DB 1; Length 550;
Best Local Similarity 18.3%; Pred. No. 8;
Matches 92; Conservative 64; Mismatches 154; Indels 192; Gaps 23;

QY 19 ESYDSAAANYMINSDMSDYL-----AVSDNFAERICSYQVPKGS 57
DB 89 ERSNFAFNCYPR--DVPDYASLRSLVAASGTFEFTGFTWGTONGSGSNACKRGP--- 143
QY 58 NCSASVSAVMSCAKODCLTLQSLKYLEAKYQPLFLPYQLYEAFFI----- 105
DB 144 -----ASGFSL---NWLTKSGSTYPV-----LNTVMPNNNDNFDKLYINGVHHPTNOQ 191
QY 106 --LFRESADNPANSTK-----RFWMKFRGRKNSYFHDLVFNLLEKNVTRDAD 152

```

```

DB 192 TNLVQASGVYVSTRSQOTIIPNIGSRWVAGSGRISIV-----TVKPPD 241
QY 153 ANDIENFASRYLYMATLYKTYTNVDEFGASFPNKLSTTGTGFGWGIKALKOIRSNLP 212
DB 242 VLIVNSGN--LIAPRGYFMRT-----GKS-----SIMSDAP 273
QY 213 LDIGTEHSVRLQHTISS--YKDYMDQIPALPKFAKRSIMV-----YORLLA 259
DB 274 IDPCISECTIPNGSIPNDKPFQVNVNKITYGACPRYVKOMTLATGMRVNPEKOTGGLFG 333
QY 260 TVAGYVDFPW---YKKWY-----MKLK-----NFMVNRVFIPTKFFN 294
DB 334 AINGFIENGEGHIDWYGFERNQSECTGQADLKSTQALIDINKLNIRVIKTEKTH 393
QY 295 --KEIREPSKALK--EKVSTDTRKDFENKIGGVDFFNKEIRDSKALKEKVSNDAD 349
DB 394 QIEKSEVEGRIQDEKVEYEDTK-----IDLMS-----YNADVLY 429
QY 350 LPENKIGCGTVDFINNEIRDPKALIRKYSTGCAEDLFENKIGGVDFINNEIRDPKAL 409
DB 430 ALEN--QHTIDLTDEMNKLFETRRLRENAEDM-----GNGCF----- 467
QY 410 IRKYTEADDLFENKIGCGTVD 431
DB 468 --KIYHKCNACIESIRNGTYD 487

RESULT 36
MAPB_MOUSE
ID MAPB_MOUSE STANDARD; PRT; 2464 AA.
AC P14873;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE [Contains: MAP1 light chain LC1].
GN MAP1B OR MTAP1B OR MTAP5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND DOMAIN
RP STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=90094539; PubMed=2480963;
RA Noble M., Lewis S.A., Cowan N.J.;
RT "The microtubule binding domain of microtubule-associated protein
RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT and tau.";
RL J. Cell Biol. 109:3367-3376(1989).
CC -I- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYOSKELETAAL CHANGES
CC THAT ACCOMPANY NERVE FIBER EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRINGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -I- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -I- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -I- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC OF MAP1B.
CC -I- SIMILARITY: TO MAP1A.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC cc entitles requires a license agreement (see http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DR DR EMBL; X51396; CAA35761.1; -.
DR PIR; S07349; QKMSPL.
DR MGD; MG1:1365778; Mcap1b.
DR InterPro; IPR000102; MAP1B_neuraxlin.
DR Pfam; PF00414; MAP1B_neuraxlin; 10.
DR PROSITE; PS00230; MAP1B_NEURAXIN; 7.
KW Microtubules; Repeat; Phosphorylation.
FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
FT REPEAT 1874 1890 MAP1B 1.
FT REPEAT 1891 1907 MAP1B 2.
FT REPEAT 1908 1924 MAP1B 3.
FT REPEAT 1925 1941 MAP1B 4.
FT REPEAT 1942 1958 MAP1B 5.
FT REPEAT 1959 1975 MAP1B 6.
FT REPEAT 1993 2009 MAP1B 7.
FT REPEAT 2010 2026 MAP1B 8.
FT REPEAT 2027 2043 MAP1B 9.
FT REPEAT 2044 2060 MAP1B 10.
FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY KKEE AND KKEIV/ REPEATS).
SQ SEQUENCE 2464 AA; 270408 MW; FBD9DD99CFBDB87 CRC64;
Query Match 4.4%; Score 103.5; DB 1; Length 2464;
Best Local Similarity 20.3%; Pred. NO. 55;
Matches 99; Conservative 77; Mismatches 180; Indels 131; Gaps 21.
OY 15 LAASSVDSAAAVYVINSMDYLSAVSNFNERICSOYPRKSNCSASVAMRCADOD 74
DB 331 LEEERSOGSTSSDMKKNLISPOLVFLNVDENLKDEP-----NIKMKSTEEA 381
OY 75 CLTLQSLKYPPEAKYOPLPLPDYOLEAFLIFKESDA-----NPANST-EKRFMR 125
DB 382 CETLLQYLN-KLSMKREPELFERSVGNITIEPY-LFQKMGCVGKLEMYLVNLPYKSKEMQYFNG 439
OY 126 FRKGNHSHFDLNVNLEKKNVTRADADTDIENFASRYLYMATLYKTYTINWDEGASFF 185
DB 440 QMTGNKKDAEILILRNGQEVDIR-----I 463
OY 186 NKLSTFTGCGFGICGRALKQIIRSNLRPLDIGNHSY--SRKHNTTSYKDYMTQIPA 241
DB 464 SYLTGVSLSLTVNHRANPRAEKITIRVLRPGN-STQYNLBLEKLNHL-----DPL----- 511
OY 242 LPKFAKRESLWVQRLLAT---VAGYVDTPWYKKWYMKLNEVNVNRFI--PTKRFENKE 296
DB 512 -----KQPLATQKDLTGQVPRPYKQ--VKLKORADSRSLKPA TKPVASKS 556
OY 297 IREPK-----ALKER--STDTKDLFEKKIKIGIYDFINKEIRDS 336
DB 557 VAKSEKEETPEVTKTSQVEKTPKVSKEKELVYKKDKRPVKTESKPSVTEKEVSKESQSPV 616
OY 337 KA-LKEKVSNDAKDLFENKIGIGIYDFINNEIR-----DPSKALIRKSTGAEDLF 386
DB 617 KAEVAEKQATESKP---KVTKDV--VKKELIKTKLEEKKEEKPKPEYVKKKDKDPIPKKD 670
OY 387 EKKIKGIGIYDFINNEIRDSKALIRKVTLEADDEENK--IQGIGYDFINKEIRDSKAL 444
DB 671 EKPRKEEVKKEIKKIKKEERKELKVEYKKEPRLDKAKKEVKKKEKVEYKKEKEPEKKE- 729
OY 445 IRKVS TE 451
DB 730 IKKISKD 736
RESULT 37
HEMA_IADM2
AC P26135. STANDARD: PRT: 566 AA.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
GN HA.
OS Influenza A virus (strain A/Duck/Memphis/928/74).
OC Viruses: ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza virus A and B group; Influenza A viruses;
OC Influenza A virus.
OX NCBI_TaxID=11367;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=92114135; Pubmed=1731092;
RA Bean W.J., Schnell M., Katz J., Kawaoaka Y., Naeye C., Gorman O.,
RA Webster R.G.;
RT "Evolution of the H3 Influenza virus hemagglutinin from human and
RT nonhuman hosts";
RL J. Virol. 66:1129-1138(1992).
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -1- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73772; -. NOT_ANNOTATED_CDS.
DR HSSP; P03437; ZHMG.
DR InterPro; IPR001364; Hemagglutn.
DR Pfam; PF00509; Hemagglutinin; 1.
DR PRINTS; PRO0329; HEMAGGLUTN12.
DR Prodom; PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; Glycoprotein; Signal.
FT SIGNAL 1 16
FT CHAIN 17 344
FT CARBOHYD 23 23 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
SO SEQUENCE 566 AA; 63572 MW; C65DFC5DE5BED80D CRC64;

Query Match 4.4%; Score 103; DB 1; Length 566;
Best Local Similarity 17.9%; Pred. No. 9.6;
Matches 92; Conservative 67; Mismatches 166; Indels 188; Gaps 22.

OY 5 DSVGVDTKTLTAAASEVSVAANAAYINSDMSDYL-----AVSD 43
Dy 93 DVFOQETMDLYVERSSARSNCPY----DVDPYASLRSLVASSGTLFETBGFVTGVTG 148
OY 44 NFAETIGQVPRKGSNCSSVSAIYMRCAKQDCITLQSLKYPELEAKYQPLTLPDPOLENA 103
Dy 149 NGGSNAACKRGP-----ASGFESRL--MWLTSGSTYPV-----LNVTMKNDNFKL 193
OY 104 FI-----LFKESDANPANSTEK-----RFMFRFRGKNHSYFHDL 138
Dy 194 YWGVNHSTQDEQTDLYVQASGRVTVSTRSQQITIIINIGSRPVPVWQSGHLSITW--- 250
OY 139 VFNLEKNTVDADATDIENFASRYLYMATLYKYKTYTNVDEFGASFENKLSFTTGLFGWG 198
Dy 251 -----TIVRPGVLYVINSNGN--LIAPRGYK----- 275
OY 199 IKRALKQIIRSNLPDIDIGTEHSVSLQHTTS--YKQVMDQIPLALPKFAKFSLMV--- 253
Dy 276 IRTGKSSIMRSDADPIDTCISIECTIPNGSIPNDKRPQVNNKITTYGACQPYVWOSTLKLATG 335

```

```

OY 254 -----VORLLATVAGYVDFPM---YKKWY-----MKLK-----NF 280
DB 336 MRVPEKRTKRGFLVATAGTENGMEGMDICWYGFRIHONSEGTGAADLSTGAALQING 395
OY 281 MVRNVEIPPKFKEIREPSKALKEKYSTDRKLFENKIGQTVDF--FNKEIRPSKA 338
DB 396 KLNVRVIEKTEKNEFHOLEKEFS-----EVEGRIGDL-ENVVEDTKIDLMSYNABL---LVA 446
OY 339 LKEKVSNDKADLFENKIGQTVDFINNEIRDPKALIRKYSTGAEDLFENKIGQTVDFI 398
DB 447 LE-----NQHITDITDSEMNKLFERTRRDLRENAEDM-----GSGCF--- 483
OY 399 NNEIRDPKALIRKYTEADDLFENKIGQTVDF 431
DB 484 -----KIYHKCDNACIESIRNGTTVD 503

RESULT 38
MYSH ACACA STANDARD: PRT: 1577 AA.
AC P47808;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE High molecular weight form of myosin I (HMMMI).
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91056121; PubMed=2243110;
RA Horowitz J.A.; Hammer J.A. III;
RT "A new Acanthamoeba myosin heavy chain. Cloning of the gene and
RL J. Biol. Chem. 265:20646-20652(1990).
CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M60954; AAA27709.1; -.
DR HSSP; P08799; 1MND.
DR InterPro: IPR000048; IO.
DR InterPro: IPR000857; MYTH4.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IO; 1.
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00784; MYTH4; 1.
DR Pfam: PF00018; SH3; 1.
DR PRINTS; PRO0193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IO; 1.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00139; MYTH4; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50096; IO; 1.
DR PROSITE; PS50096; IO; 1.
KW Myosin; Actin-binding; ATP-binding; Methylation; Multigene family;
KM SH3 domain.
FT DOMAIN 1 754 MYOSIN HEAD-LIKE.
FT DOMAIN 755 782 IO.
FT DOMAIN 1519 1577 SH3.
FT NP_BIND 168 175 ATP (POTENTIAL).
FT DOMAIN 628 650 ACTIN-BINDING (POTENTIAL).

```

```

FT DOMAIN 730 744 ACTIN-BINDING (POTENTIAL).
FT DOMAIN 277 280 POLY-GLY.
FT DOMAIN 1471 1476 POLY-ALA.
FT DOMAIN 1482 1489 POLY-GLN.
FT DOMAIN 1493 1496 POLY-GLN.
FT DOMAIN 1506 1511 POLY-GLN.
FT MOD_RES 119 119 METHYLATION (TRI-) (POTENTIAL).
SQ SEQUENCE 1577 AA; 177529 MW; BAF28EBA8B7438C CRC64;

Query Match 4.4%; Score 103; DB 1; Length 1577;
Best Local Similarity 20.0%; Pred. No. 34;
Matches 98; Conservative 78; Mismatches 185; Indels 130; Gaps 21;

OY 6 SVGDVTKTLTAASEVSDSANAAMINSDMDLYLSAINDNAEICQVPRGSCMSVSA 65
DB 191 STNGVEONILLANPILFESFNMAKTLRNSSRFKWEIHPDQ-----KGSICGASTIN 243
OY 66 YMSRCAKODCLTLOSILKYPLEAKYOPTLPDPYOLEAFTLFESDANPANSTERKFRMR 125
DB 244 HL-----LEKSRVYQIKGERN-----PRIVATELY---KAP-----R 274
OY 126 FRGKNHSTFFHDLVFNILKKNVTRDADAT-DIENPASRYLYMATLYKTYTNVDFGASF 184
DB 275 SRGGCGSSPARPESFFFLSQSCIDVEGVDFKFEFEVYLCHQARVRVQFSEDDIN--- 331
OY 185 FNKLSTTGIFGWC-----IKRALKQIIRSLPLDIGT--EHSVSRID 225
DB 332 -NCMELISATLHIGNEFVSGGKNVETSVANREEVKIVATLKVDPALDEONVTSKLM 390
OY 226 HTSSYKYDWDQIPALPKAKRFSLMVVORLLATVAGYVDFPMYKWKMLKNFVNRV 285
DB 391 EI-----KGDPTRIPLP-----VQATDATNA-----LAKAITSKLPDWLVKTI 430
OY 286 FIFTKFFPKKEIREPSKALKEKYSTDRKLF-----ENKIGQTVDFENKIRDP----- 335
DB 431 -----NESMEPQKGAT-TTIGVLDIFGEIPIDKNSFEOLCINFNEKIQOHFNQY 480
OY 336 SKALKEKVSNDKADLFENKIGQTVDFINNEIRDPKALIRKYSTGAEDLFENKIGQTV 395
DB 481 TFLLEKLYOSEEVKTEH-----ITFIDN-----PYLDLEKKQPGOLM 530
OY 396 DFINNEIRPSKA---LIRKYTEADDLFENKIGQTVDFINKE-----IRDPKALI 445
DB 521 LVVDEQISIPKSSDANFFIKANQTAARSTQLNGGEDSPDFIHKYAGVIVDSGMLE 580
OY 446 RKVSTEARNDLL 456
DB 581 KNKDTLQKDLL 591

RESULT 39
TOP2_SCHPO STANDARD: PRT: 1485 AA.
AC P08096; O74336;
DT 01-AUG-1988 (Rel. 08, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA topoisomerase II (EC 5.99.1.3).
GN TOP2 OR SPBC1A4.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87053875; PubMed=3023070;
RA Demura T.; Morikawa K.; Yanagida M.;
RT "The nucleotide sequence of the fission yeast DNA topoisomerase II
RT gene: structural and functional relationships to other DNA
RL EMBO J. 5:2355-2361(1986).
RN [2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Volckaert G.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP DOMAINS.
RA MEDLINE-92049339; PubMed-1658625;
RA Shiozaki K., Yanagida M.;
RT "A functional 125-kDa core polypeptide of fission yeast DNA
RT topoisomerase II.";
RL Mol. Cell. Biol. 11:6093-6102(1991).
RN [4]
RP DOMAINS, AND PHOSPHORYLATION.
RA MEDLINE-93077665; PubMed-1332977;
RA Shiozaki K., Yanagida M.;
RT "Functional dissection of the phosphorylated terminl of fission yeast
RT DNA topoisomerase II.";
RL J. Cell Biol. 119:1023-1036(1992).
CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
CC MAKES DOUBLE-STRAND BREAKS.
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES AT BOTH EXTREMITIES OF THE
CC PROTEIN.
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: X04326; CAA27857.1; ALT_INIT.
DR EMBL: AL031174; CAA20107.1; -.
DR PIR: A24897; ISZPT2.
DR HSP: P06786; I8GM.
DR InterPro: IPR003957; CBFA_NFYB_topis.
DR InterPro: IPR001241; DNA_topoisoi.
DR InterPro: IPR002205; DNA_topoisoi.
DR InterPro: IPR003594; HATPase_c.
DR Pfam: PF00204; DNA_topoisoi.1.
DR Pfam: PF00521; DNA_topoisoi.1.
DR Pfam: PF02518; HATPase_c.1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PRINTS: PR00418; TP12FAMILY.
DR PRODom: PD000616; DNA_topoisoi.1.
DR SMART: SM00387; HATPase_c.1.
DR SMART: SM00433; TOP2c.1.
DR SMART: SM00434; TOP2c.1.
DR PROSITE: PS00177; TOPOISOMERASE_II.1.
KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Phosphorylation;
KW Nuclear protein.
FT NP_BIND 206 211 ATP (POTENTIAL).
FT ACT_SITE 835 835 DNA CLEAVAGE (BY SIMILARITY).
FT CONFLICT 994 994 N -> I (IN REF. 1).
SQ SEQUENCE 1485 AA; 167891 MW; 6D88F76243361B2F CRC64;

```

Query Match 4.3%; Score 102.5; DB 1; Length 1485;
 Best Local Similarity 21.0%; Pred. No. 34;
 Matches 113; Conservative 62; Mismatches 157; Indels 207; Gaps 30;

QY 30 INSDMSDYLSAVSDNFAIRICQVYKSGNSCASVAYMRCAGK-----QDCLT--LQS 80
 Db 3 IDADSDVEDEASGD--ENVLPNTT--TKRAKSTSSSKRAKKASTPDLRQSLSLMSMAS 58

```

QY 81 LKYP-----EAKYPLT-----LPDYQLEAFILFESDANPANSTERR 122
Db 59 EQPLVLTNGNSNVSYQYRLTPREHVLRRPDY-----IGSIEPTSEM 105
QY 123 WMRFRGRK-----HSYFHDVFNLLK-----NTRRODAT----- 154
Db 106 WV-FDSEKKNKLDYKAVTYVPGLYKIFDELIVAAONKVRDPMNTLKYTLDEAVVISIT 164
QY 155 -----DIENFASRYLYMATLY-----KTYTNDEFGASEFNKLSFTTGLFGIGIRALKQ 205
Db 165 NNGKQPIPIFHDKEKIIYIPELIFGNLLTSSNYD-----NOKKVTGGANGYGAK--LCN 216
QY 206 IIRSNLPDITGTHSVSLQ-----ITSSKYDMYDQIPALPKFARFSL- 251
Db 217 IFTFEVETADKERNKKYKQTYWDMNSRKSSEPVITSLKPKDEYKRTIKFKPLAK-FCGD 275
QY 252 -----HWQVRLATVAGYVDTTPYKKKMYMLKFMNRRVFIPIKKFENK---IREPS 301
Db 276 KIDDMVSTIKRRIYDMGTVR-----ERKVLNRRISISGPK 314
QY 302 KALKRKVSTDTKDLFENKIGQGTVDFFNKELRDPKALKRKVSNDAKDL-FENKIGQ-GT 359
Db 315 KYEMKYLASDT-----KPDPEPRVYIEHV-NDRMVAFAVSDGQFKQ 356
QY 360 VDFINNEIRDPKALIRKVTGAEGLFENKIGQGT-VDFINNEIRDPKALIRKYVTEAD 418
Db 357 VSFVNN-----IST-----IRGTHVNVANKIYDAIDEVVK----- 389
QY 419 DLFENK-----IGQGVDFINKEIRDP-----KALIRKVS-----TEADNILE 457
Db 390 ---ENKKAQVKAFOIKRYVQVFNQCIENPSFDSOTKETLTTKVSAFSGSOCLSLSKFLK 445

```

```

RESULT 40
LMA2_MOUSE
ID LMA2_MOUSE STANDARD: PRT: 3106 AA.
AC Q60675; Q05003; Q64061.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-2 chain precursor (Laminln M chain) (Merosin heavy
DE chain).
GN LMA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Embryo, and Heart;
RX MEDLINE-95316259; PubMed-7795883;
RA Bernier S.M., Utani A., Sugiyama S., Dol T., Pollastina C.,
RA Yamada Y.;
RT "Cloning and expression of laminln alpha 2 chain (M-chain) in the
RT mouse.";
RL Molec Biol. 14:447-455(1995).
RN [2]
RP SEQUENCE OF 2162-2279 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RX MEDLINE-93346725; PubMed-8345183;
RA Chang A.C., Wadsworth S., Colligan J.E.;
RT "Expression of merosin in the thymus and its interaction with
RT thymocytes.";
RL J. Immunol. 151:1789-1801(1993).
RN [3]
RP SEQUENCE OF 64-281 FROM N.A.
RX MEDLINE-95179178; PubMed-7874173;
RA Xu H., Wu X.R., Wewer U.M., Engvall E.;
RT "Murine muscular dystrophy caused by a mutation in the laminln alpha
RT 2 (Lama2) gene.";
RL Nat. Genet. 8:297-302(1994).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 2932-3106.

```

RX MEDLINE-20085745; PubMed-10619025;
 RA Hohenester E., Tisi D., Tails J.F., Timpl R.;
 RT "The crystal structure of a laminin G-like module reveals the
 RT molecular basis of alpha-dystroglycan binding to laminins, perlecan,
 RT and agrin.";
 RL Mol. Cell 4:783-792(1999).
 CC -1- FUNCTION: BINDING TO CELLS VIA A HIGH AFFINITY RECEPTOR, LAMININ
 CC IS THOUGHT TO MEDIATE THE ATTACHMENT, MIGRATION, & ORGANIZATION OF
 CC CELLS INTO TISSUES DURING EMBRYONIC DEVELOPMENT BY INTERACTING
 CC WITH OTHER EXTRACELLULAR MATRIX COMPONENTS.
 CC -1- SUBUNIT: LAMININ IS A COMPLEX GLYCOPROTEIN, CONSISTING OF THREE
 CC DIFFERENT POLYPEPTIDE CHAINS (ALPHA, BETA, GAMMA), WHICH ARE BOUND
 CC TO EACH OTHER BY DISULFIDE BONDS INTO A CROSS-SHAPED MOLECULE
 CC COMPRISING ONE LONG & THREE SHORT ARMS WITH GLOBULES AT EACH END.
 CC THE ALPHA-2 CHAIN IS A SUBUNIT OF LAMININ-2 (MEROSIN) AND LAMININ-
 CC 4 (S-MEROSIN).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
 CC -1- DISEASE: DEFECTS IN LAMA2 ARE A CAUSE OF MORINE MUSCULAR DYSTROPHY
 CC (DY2J).
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: U12147; AAC52165.1; -;
 DR EMBL: X69869; CA449502.1; -;
 DR EMBL: S75315; AAB33573.1; -;
 DR PDB: 1000; 03-DEC-99.
 DR MGD: MGI:99912; Lama2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001886; LAMNT.
 DR InterPro: IPR000034; Laminin_B.
 DR InterPro: IPR002049; Laminin_EGF.
 DR Pfam: PF00052; Laminin_B; 2.
 DR Pfam: PF00053; Laminin_EGF; 15.
 DR Pfam: PF00054; Laminin_G; 5.
 DR Pfam: PF00055; Laminin_Nterm; 1.
 DR PRINTS: PR00011; EGF/LAMININ.
 DR ProDom: PD002082; LAMNT; 1.
 DR ProDom: PD003031; Laminin_B; 2.
 DR SMART: SM00180; EGF_Lam; 15.
 DR SMART: SM00001; EGF-like; 1.
 DR SMART: SM00281; Lamb; 2.
 DR SMART: SM00282; Lamb; 5.
 DR SMART: SM00136; LAMNT; 1.
 DR PROSITE: PS00022; EGF_1; 11.
 DR PROSITE: PS01186; EGF_2; 3.
 DR PROSITE: PS01248; LAMININ TYPE EGF; 14.
 DR PROSITE: PS50025; LAM_G_DOMAIN; 5.
 KM Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KM Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 KW SIGNAL
 FT CHAIN 1 22
 FT CHAIN 23 3106 LAMININ ALPHA-2 CHAIN.
 FT CHAIN 23 282 LAMININ N-TERMINAL (DOMAIN VI).
 FT CHAIN 283 339 LAMININ EGF-LIKE 1.
 FT CHAIN 340 409 LAMININ EGF-LIKE 2.
 FT CHAIN 410 464 LAMININ EGF-LIKE 3.
 FT CHAIN 465 513 LAMININ EGF-LIKE 4.
 FT CHAIN 514 523 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 524 719 LAMININ DOMAIN IV 1 (DOMAIN IV B).
 FT DOMAIN 752 752 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 753 802 LAMININ EGF-LIKE 6.
 FT DOMAIN 803 860 LAMININ EGF-LIKE 7.
 FT DOMAIN 861 913 LAMININ EGF-LIKE 8.
 FT DOMAIN 914 962 LAMININ EGF-LIKE 9.
 FT DOMAIN 963 1009 LAMININ EGF-LIKE 10.
 FT DOMAIN 1010 1055 LAMININ EGF-LIKE 11.
 FT DOMAIN 1056 1101 LAMININ EGF-LIKE 12.
 FT DOMAIN 1102 1161 LAMININ EGF-LIKE 13.
 FT DOMAIN 1162 1171 LAMININ EGF-LIKE 14 (N-TERMINAL).
 FT DOMAIN 1172 1375 LAMININ DOMAIN IV 2 (DOMAIN IV A).
 FT DOMAIN 1376 1415 LAMININ EGF-LIKE 1.
 FT DOMAIN 1416 1464 LAMININ EGF-LIKE 15 (C-TERMINAL).
 FT DOMAIN 1465 1522 LAMININ EGF-LIKE 16.
 FT DOMAIN 1523 1569 LAMININ EGF-LIKE 17.
 FT DOMAIN 1570 2140 LAMININ II AND I.
 FT DOMAIN 2141 2324 LAMININ G-LIKE 1.
 FT DOMAIN 2326 2517 LAMININ G-LIKE 2.
 FT DOMAIN 2522 2706 LAMININ G-LIKE 3.
 FT DOMAIN 2759 2930 LAMININ G-LIKE 4.
 FT DOMAIN 2929 3106 LAMININ G-LIKE 5.
 FT DOMAIN 1662 1863 LAMININ G-LIKE 6.
 FT DOMAIN 1923 2146 COILED COIL (POTENTIAL).
 FT DOMAIN 283 292 COILED COIL (POTENTIAL).
 FT DISULFID 285 303 BY SIMILARITY.
 FT DISULFID 305 314 BY SIMILARITY.
 FT DISULFID 317 337 BY SIMILARITY.
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 374 BY SIMILARITY.
 FT DISULFID 377 386 BY SIMILARITY.
 FT DISULFID 389 407 BY SIMILARITY.
 FT DISULFID 410 422 BY SIMILARITY.
 FT DISULFID 412 438 BY SIMILARITY.
 FT DISULFID 440 449 BY SIMILARITY.
 FT DISULFID 452 462 BY SIMILARITY.
 FT DISULFID 465 478 BY SIMILARITY.
 FT DISULFID 467 482 BY SIMILARITY.
 FT DISULFID 484 493 BY SIMILARITY.
 FT DISULFID 496 511 BY SIMILARITY.
 FT DISULFID 511 511 BY SIMILARITY.
 FT DISULFID 511 762 BY SIMILARITY.
 FT DISULFID 755 769 BY SIMILARITY.
 FT DISULFID 772 781 BY SIMILARITY.
 FT DISULFID 784 800 BY SIMILARITY.
 FT DISULFID 803 818 BY SIMILARITY.
 FT DISULFID 805 828 BY SIMILARITY.
 FT DISULFID 831 840 BY SIMILARITY.
 FT DISULFID 843 858 BY SIMILARITY.
 FT DISULFID 861 875 BY SIMILARITY.
 FT DISULFID 863 882 BY SIMILARITY.
 FT DISULFID 885 894 BY SIMILARITY.
 FT DISULFID 897 911 BY SIMILARITY.
 FT DISULFID 914 926 BY SIMILARITY.
 FT DISULFID 916 933 BY SIMILARITY.
 FT DISULFID 935 944 BY SIMILARITY.
 FT DISULFID 947 960 BY SIMILARITY.
 FT DISULFID 963 975 BY SIMILARITY.
 FT DISULFID 965 981 BY SIMILARITY.
 FT DISULFID 983 992 BY SIMILARITY.
 FT DISULFID 995 1007 BY SIMILARITY.
 FT DISULFID 1010 1019 BY SIMILARITY.
 FT DISULFID 1012 1026 BY SIMILARITY.
 FT DISULFID 1028 1037 BY SIMILARITY.
 FT DISULFID 1040 1053 BY SIMILARITY.
 FT DISULFID 1056 1068 BY SIMILARITY.
 FT DISULFID 1058 1075 BY SIMILARITY.
 FT DISULFID 1077 1089 BY SIMILARITY.
 FT DISULFID 1089 1096 BY SIMILARITY.
 FT DISULFID 1416 1425 BY SIMILARITY.
 FT DISULFID 1418 1432 BY SIMILARITY.
 FT DISULFID 1435 1444 BY SIMILARITY.
 FT DISULFID 1447 1462 BY SIMILARITY.
 FT DISULFID 1465 1480 BY SIMILARITY.


```

FT DISULEID 1467 1490 BY SIMILARITY.
FT DISULEID 1493 1502 BY SIMILARITY.
FT DISULEID 1505 1520 BY SIMILARITY.
FT DISULEID 1523 1535 BY SIMILARITY.
FT DISULEID 1525 1542 BY SIMILARITY.
FT DISULEID 1544 1553 BY SIMILARITY.
FT DISULEID 1556 1567 BY SIMILARITY.
FT DISULEID 1570 1570 INTERCHAIN (PROBABLE).
FT DISULEID 1574 1574 INTERCHAIN (PROBABLE).
FT CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

Query Match 4.38; Score 102.5; DB 1; Length 3106;
 Best Local Similarity 20.38; Pred. No. 85;
 Matches 109; Conservative 68; Mismatches 178; Indels 183; Gaps 26;

```

QY 4 SDSYGVYTKTLASSESANANWYMSDYLAVSDNFAIRICQVPGKSGCSASY 63
Db 1671 AESLEEFIKGLVQDAEINE--KAVKLNETIGN-----QDKTAERNLELOK-----EI 1717
QY 64 SAYMSRCAKODCLTLQSLKYPLEAKYQPLTPPYOLEAFILFKESDANPANSTERKRW 123
Db 1718 DRMLKELRSKDLQT-----OKEVA-----EDELVAAGELKRVN 1751
QY 124 MRF--RGKHNHYPHDLVFNLE-KNVTROA-----DATDIENFASRYLYMATLYKTYT 175
Db 1752 KLFGEPPRAQNDMEKDLQOKLAEYKNNKLDAMDLLRENTDTRDANR--LSAANQKNMT 1808
QY 176 NVDEFGASFEFNKLSFTTGLFGMGIKRALKOIIR-SNLPDIDGTEHSVRLQHTSSYKDY 234
Db 1809 ILETKKKAI-----EGSKROIENTLKEGNDILD-----EANGLGEINSVIDY 1851
QY 235 MDQIPLPKFAKRFSLMV-----VQRL-----ATVAGYVDTFW 269
Db 1852 VDDIKTKLPMSSELSDKIDDLAQEIKDRRLAEKVFQASHAAQLNDSSAVLDGILD--- 1908
QY 270 YKKWYMKLKNMNV-----RVFIPTKKFENKEIREPSKALKEKYSTDPKDLFENKIGGT 324
Db 1909 -----EKKNISFNATAAFRAVSNIKDYIDE-----AEKVAREAKEL-----AOGA 1948
QY 325 VDFNKEIRDPKSKALKE-----KVSNDK-----DLFENKIGGTVDFI 363
Db 1949 -----TKLATSPQGLLKEBDAKSLQKSPRLNEAKKLANLVKGNHNDLNDLTKRLETADLR 2004
QY 364 N-----NEIRDPKSKALIRKYSTGAEDLFENKIGGTVDFINNEIRDPKSKALIRKYTE 416
Db 2005 NSGLGLANDTMDKLSAITNDTAAKLAQAVKEKA-----REANDTAKAVLAQV--- 2051
QY 417 ADDLFENKIG-----QGTVDFINKEIRDPK-----ALIRKYSTEADNLEK 458
Db 2052 -KDLHQNDLQKQNYKNKLADSVAKTNNAVVDPSKNIITADAGTSVKNLEQEDRLLDK 2108

```

Search completed: October 19, 2002, 07:49:52
 Job time : 36 secs